

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:15:33 ; Search time 9.71143 Seconds
(without alignments)
2813.016 Million cell updates/sec

Title: US-10-006-011A-9

Perfect score: 1566

Sequence: 1 CERQPCQHGATCMAGEYEF.....QSLDLQHRQAQANTRPQPS 284

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1566	100.0	4391	2 A38096	perlecan precursor
2	1370.5	87.5	3707	2 S1852	heparan sulfate pr
3	349	22.3	3375	2 T19821	hypothetical prote
4	333.5	21.3	1959	1 AGRT	agrin - rat
5	314.5	20.1	1328	2 T43060	agrin - electric r
6	301.5	19.3	1955	1 AGCH	agrin precursor -
7	283.5	18.1	1531	2 T42218	slit-1 protein hom
8	264	16.9	4307	2 T20721	hypothetical prote
9	264	16.9	5147	1 IJFTM	cadherin-related t
10	249	15.9	601	2 T22025	hypothetical prote
11	249	15.9	601	2 D89711	protein F40E10.4 [
12	246	15.7	1523	2 T13953	MEG5 protein - ra
13	243.5	15.3	1025	2 T42626	secreted leucine-r
14	223.5	14.3	1751	1 MWHUMH	laminin alpha-2 ch
15	217	13.9	1438	2 A48216	neurexin III-alpha
16	217	13.9	1471	2 B48218	neurexin III-alpha
17	217	13.9	1578	2 T48216	G-cadherin - sea u
18	214	13.7	2809	2 T30213	laminin alpha-2 ch
19	207.5	13.3	3106	1 S53668	notch3 protein - h
20	199	12.7	2321	2 S78549	neurexin I-alpha
21	197.5	12.6	1530	2 T45944	hypothetical prote
22	197.5	12.5	2610	2 T20368	neurexin I-alpha p
23	196.5	12.5	1507	2 A40228	neurexin II-alpha
24	194.5	12.4	1715	2 C40228	seven-pass transme
25	191	12.2	3034	2 T14119	laminin alpha-1 ch
26	188.5	12.0	3075	2 S14458	hypothetical prote
27	185.5	11.8	3672	2 T23433	probable laminin a
28	185.5	11.8	3704	2 T37316	DN-cadherin - fru
29	184	11.7	3097	2 T00021	

RESULT 1

A38096

perlecan precursor - human

N;Alternate names: basement

C;Species: Homo sapiens (man)

C;Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #test change 05-Nov-1999

C;Accession: A38096, S19256, S77946; A41059; A40306; B33625, A33625; A41736

R;Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.

J. Biol. Chem. 267, 8544-8557, 1992

A;Title: Primary structure of the human heparan sulfate proteoglycan from basement membr

tor, laminin, neural cell adhesion molecules, and epidermal growth factor.

A;Reference number: A38096; MUID:92235084; PMID:1569102

A;Accession: A38096

A;Molecule type: mRNA

A;Residues: 1-4391 <MUI>

A;Cross-references: GB:M8289; NID:9184426; PIDN:AAAS2700.1; PID:9184427

R;Kallunki, P.; Tryggvason, K.

J. Cell Biol. 116, 559-571, 1992

A;Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD prot

ell adhesion molecules, and epidermal growth factor.

A;Reference number: A41736; MUID:92112994; PMID:1730768

A;Accession: S19256

A;Molecule type: mRNA

A;Residues: 1-57, 'D', 59-434, 'A', 436, 'FL', 438-449, 'O', 451-502, 'A', 503-792, 'K', 794-908, 'R',

71-2979, 'H', 2981-2994, 'G', 2996-3167, 'T', 3169-3240, 'R', 3242-3426, 'R', 3428-3631, 'Q', 3633-36

A;Cross-references: EMBL:X62515; NID:929469; PIDN:CAA4373.1; PID:929470

R;Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestila, M.; Shows, T.B.; Tryggvason, K.

Genomics 11, 389-396, 1991

A;Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the ge

A;Reference number: A41059; MUID:92120660; PMID:1685141

A;Accession: A41059

A;Molecule type: mRNA

A;Residues: 1-57, 'D', 59-434, 'A', 436, 'FL', 438-449, 'O', 451-502, 'A', 503-792, 'K', 794-908, 'R',

71-2979, 'H', 2981-2994, 'G', 2996-3167, 'T', 3169-3240, 'R', 3242-3426, 'R', 3428-3631, 'Q', 3633-40

A;Cross-references: EMBL:X62515; NID:929469; PIDN:CAA4373.1; PID:929470

R;Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestila, M.; Shows, T.B.; Tryggvason, K.

Genomics 11, 389-396, 1991

A;Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the ge

A;Reference number: A41059; MUID:92120660; PMID:1685141

A;Accession: A41059

A;Molecule type: mRNA

A;Residues: 1-57, 'D', 59-434, 'A', 436, 'FL', 438-449, 'O', 451-502, 'A', 503-792, 'K', 794-908, 'R',

71-2979, 'H', 2981-2994, 'G', 2996-3167, 'T', 3169-3240, 'R', 3242-3426, 'R', 3428-3631, 'Q', 3633-40

A;Cross-references: EMBL:X62515; NID:929469; PIDN:CAA4373.1; PID:929470

R;Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestila, M.; Shows, T.B.; Tryggvason, K.

Genomics 11, 389-396, 1991

A;Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the ge

A;Reference number: A41059; MUID:92120660; PMID:1685141

A;Accession: A41059

A;Molecule type: mRNA

A;Residues: 1-57, 'D', 59-434, 'A', 436, 'FL', 438-449, 'O', 451-502, 'A', 503-792, 'K', 794-908, 'R',

71-2979, 'H', 2981-2994, 'G', 2996-3167, 'T', 3169-3240, 'R', 3242-3426, 'R', 3428-3631, 'Q', 3633-40

A;Cross-references: EMBL:X62515; NID:929469; PIDN:CAA4373.1; PID:929470

R;Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestila, M.; Shows, T.B.; Tryggvason, K.

Genomics 11, 389-396, 1991

A;Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the ge

A;Reference number: A41059; MUID:92120660; PMID:1685141

A;Accession: A41059

A;Molecule type: mRNA

A;Residues: 1-57, 'D', 59-434, 'A', 436, 'FL', 438-449, 'O', 451-502, 'A', 503-792, 'K', 794-908, 'R',

A;Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclonal antibodies.
 A;Reference number: A33625; MUID:90078352; PMID:2687294
 A;Accession: B33625
 A;Molecule type: protein
 A;Residues: 1379-1384,'X',1386-1388,'X',1390-1398 <HE2>
 A;Accession: A33625
 A;Molecule type: protein
 A;Residues: 2166-2171,'X',2173-2175,'X',2177-2185 <HE3>
 A;Note: peptide potentially matches four different regions of sequence shown
 C;Genetics:
 A;Gene: GDB:HSPG2
 A;Cross-references: GDB:126372; OMIM:142461
 A;Map position: lp36.1-lp36.1
 C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repeat
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-4391/Product: perlecan #status predicted <MAT>
 F;22-193/Domain: I <DOM1>
 F;194-530/Domain: II <DOM2>
 F;199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F;285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F;325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F;368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
 F;531-1676/Domain: III <DOM3>
 F;1159-1206/Domain: laminin-type EGF-like homology <LEG>
 F;1563-1610/Domain: laminin-type EGF-like homology <EGV>
 F;1613-1668/Domain: laminin-type EGF-like homology <LEG8>
 F;1677-3686/Domain: IV <DOM4>
 F;2007-2034/Domain: transmembrane #status predicted <TRM>
 F;3687-4391/Domain: V <DOM5>
 F;3845-3880/Domain: EGF homology <EGF1>
 F;3888-3921/Domain: EGF homology <EGF>
 F;3953-4106/Domain: laminin G repeat homology <LG2>
 F;4147-4175/Domain: EGF homology <EGF2>
 F;4149-4151/Region: motor neuron attachment (L-R-E) motif
 F;4299-4301/Region: motor neuron attachment (L-R-E) motif
 F;65.71,76/Binding site: heparan sulfate (Ser) (covalent) #status predicted
 F;89.554,1755,2121,3072,3105,3279,3780,3836,4068/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;2995,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 1566; DB 2; Length 4391;

Best Local Similarity 100.0%; Pred. No. 3.2e+105;

Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CERQPCQHGATCMPAGEYEFQCLCRDGFKGLCEHENPCOLRBPCLHGCTCGTRCLCL	60
DB	4108	CERQPCQHGATCMPAGEYEFQCLCRDGFKGLCEHENPCOLRBPCLHGCTCGTRCLCL	4167
QY	61	PGFSGPRCQCGHGIAPSDWHLGSGGNDAPQGYGAYFHDGDFLAPFGHVSRLPEVP	120
DB	4168	PGFSGPRCQCGHGIAPSDWHLGSGGNDAPQGYGAYFHDGDFLAPFGHVSRLPEVP	4227
QY	121	ETIELEVRTSTASGLLLWQGVGEAGQKDFISLQDQGHVFRYOLGSGEARLVSEDP	180
DB	4228	ETIELEVRTSTASGLLLWQGVGEAGQKDFISLQDQGHVFRYOLGSGEARLVSEDP	4287
QY	181	INDGEWHRVTLRGRGSIQVDGEELVSGRSPGNVAVNAKGSVYTGAPDVATLTGGR	240
DB	4288	INDGEWHRVTLRGRGSIQVDGEELVSGRSPGNVAVNAKGSVYTGAPDVATLTGGR	4347
QY	241	FSSGITGCVKNLVLSARPAPPPQPLDLOHRAQAGANTRPCPS	284
DB	4348	FSSGITGCVKNLVLSARPAPPPQPLDLOHRAQAGANTRPCPS	4391

RESULT 2

S18252

heparan sulfate proteoglycan - mouse

N;Alternate names: perlecan

C;Species: Mus musculus (house mouse)

C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999

C;Accession: S18252; A31917; B31917; S66460

R;Noonan, D.M.; Full, A.; Valente, P.; Cai, S.; Horigan, E.; Sasaki, M.; Yamada, Y.; He
 J. Biol. Chem. 266, 22939-22947, 1991
 A;Title: The complete sequence of perlecan, a basement membrane heparan sulfate proteogly
 adhesion molecule.
 A;Reference number: S18252; MUID:92078153; PMID:1744087
 A;Accession: S18252
 A;Molecule type: mRNA
 A;Residues: 1-3707 <NOO>
 A;Cross-references: EMBL:M71174; NID:G200295; PIDN:AAA39911.1; PID:G200296
 R;Noonan, D.M.; Horigan, E.A.; Ledbetter, S.R.; Vogeli, G.; Sasaki, M.; Yamada, Y.; Hase
 J. Biol. Chem. 263, 16379-16387, 1988
 A;Title: Identification of cDNA clones encoding different domains of the basement membra
 A;Reference number: A92680; MUID:89034110; PMID:2372708
 A;Accession: A31917
 A;Molecule type: mRNA
 A;Residues: 940-1601 <NO2>
 A;Cross-references: GB:J04054; NID:G200252; PIDN:AAA39899.1; PID:G200253
 A;Accession: B31917
 A;Molecule type: mRNA
 A;Residues: 1870-2600 <NO3>
 A;Cross-references: GB:J04055; NID:G200300; PIDN:AAA39912.1; PID:G200301
 R;Schulze, B.; Mann, K.; Battistutta, R.; Wiedemann, H.; Timpl, R.
 Eur. J. Biochem. 231, 551-556, 1995
 A;Title: Structural properties of recombinant domain III-3 of perlecan containing a glob
 A;Reference number: S66460; MUID:95377282; PMID:7649154
 A;Accession: S66460
 A;Molecule type: protein
 A;Residues: 1272-1274,'X',1276,'X',1278-1279 <SCH>
 C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
 C;Keywords: glycoprotein

F;199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>	
F;285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>	
F;325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>	
F;368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>	
F;764-811/Domain: laminin-type EGF-like homology <LEG>	
F;1159-1206/Domain: laminin-type EGF-like homology <LEG7>	
F;1563-1610/Domain: laminin-type EGF-like homology <EGV>	
F;1613-1668/Domain: laminin-type EGF-like homology <LEG8>	
F;3163-3198/Domain: EGF homology <EGF>	
F;3270-3423/Domain: laminin G repeat homology <LG2>	
F;3464-3492/Domain: EGF homology <EGF7>	
F;1256,1891,2336,2394,2427/Binding site: carbohydrate (Asn) (covalent) #status predicted	

Query Match 87.5%; Score 1370.5; DB 2; Length 3707;

Best Local Similarity 86.3%; Pred. No. 4e+91;

Matches 245; Conservative 20; Mismatches 18; Indels 1; Gaps 1;

QY	1	CERQPCQHGATCMPAGEYEFQCLCRDGFKGLCEHENPCOLRBPCLHGCTCGTRCLCL	60
DB	3425	CERQPCQHGATCMPAGEYEFQCLCRDGFKGLCEHENPCOLRBPCLHGCTCGTRCLCL	3484
QY	61	PGFSGPRCQCGHGIAPSDWHLGSGGNDAPQGYGAYFHDGDFLAPFGHVSRLPEVP	120
DB	3485	PGFSGPRCQCGHGIAPSDWHLGSGGNDAPQGYGAYFHDGDFLAPFGHVSRLPEVP	3544
QY	121	ETIELEVRTSTASGLLLWQGVGEAGQKDFISLQDQGHVFRYOLGSGEARLVSEDP	180
DB	3545	ETIELEVRTSTADGLLLWQGV-VREASRSDKDFISLQDQGHVFRYOLGSGEARLVSEDP	3603
QY	181	INDGEWHRVTLRGRGSIQVDGEELVSGRSPGNVAVNAKGSVYTGAPDVATLTGGR	240
DB	3604	INDGEWHRVTLRGRGSIQVDGEELVSGRSPGNVAVNAKGSVYTGAPDVATLTGGR	3663
QY	241	FSSGITGCVKNLVLSARPAPPPQPLDLOHRAQAGANTRPCPS	284
DB	3664	FSSGITGCVKNLVLSARPAPPPQPLDLOHRAQAGANTRPCPS	3707

RESULT 3

T19821

hypothetical protein ZC101.2e - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 23-Sep-2002

C;Accession: T19821, T19819, T19820, T27490, T27488, T27489, T27487, A47648, B47648; C47
R;Baynes, C.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19182
A;Accession: T19821
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3375 <W1>
A;Cross-references: EMBL:Z93375; PIDN:CAB07569.1; GSPDB:GN00020; CESP:ZC101.2e
A;Experimental source: clone C38C
A;Accession: T19819
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2441, 'R', 'KRKH', 3369, 'GN', 3372-3373, 'G', 3375, 'RLRHRRNAQNGPL
A;Cross-references: EMBL:Z93375; PIDN:CAB07567.1; GSPDB:GN00020; CESP:ZC101.2a
A;Experimental source: clone C38C
A;Accession: T19820
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1694, 'H', 1883-2441, 'R', 'KRKH', 3369, 'GN', 3372-3373, 'G', 3375, 'RLRHRRNAQNGPL
A;Cross-references: EMBL:Z93375; PIDN:CAB07568.1; GSPDB:GN00020; CESP:ZC101.2c
A;Experimental source: clone C38C
R;Peroy, C.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z20375
A;Accession: T27490
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3375 <W1>
A;Cross-references: EMBL:Z93395; PIDN:CAB07708.1; GSPDB:GN00020; CESP:ZC101.2e
A;Experimental source: clone ZC101
A;Accession: T27488
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2441, 'R', 'KRKH', 3369, 'GN', 3372-3373, 'G', 3375, 'RLRHRRNAQNGPLSKRTTTTKLFGSW
A;Cross-references: EMBL:Z93395; PIDN:CAB07706.1; GSPDB:GN00020; CESP:ZC101.2a
A;Experimental source: clone ZC101
A;Accession: T27489
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1694, 'H', 1883-2441, 'R', 'KRKH', 3369, 'GN', 3372-3373, 'G', 3375, 'RLRHRRNAQNGPL
A;Cross-references: EMBL:Z93395; PIDN:CAB07707.1; GSPDB:GN00020; CESP:ZC101.2c
A;Experimental source: clone ZC101
A;Accession: T27487
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1128, 1290, 'DFARNSPS', 1299, 'NSS', 1303-1304, 'R', 'RHR', 1544-1545, 'RIRVRS', 155
A;Cross-references: EMBL:Z93395; PIDN:CAB07704.1; GSPDB:GN00020; CESP:ZC101.2b
A;Experimental source: clone ZC101
R;Rogalski, T.M.; Williams, B.D.; Mullen, G.P.; Moerman, D.G.
Genes Dev. 7, 1471-1484, 1993
A;Title: Products of the unc-52 gene in *Caenorhabditis elegans* are homologous to the *ced*
A;Reference number: A47648; MUID:93339574; PMID:8393416
A;Accession: A47648
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-546, 'P', 548-2441, 'R', 'KRKH', 3369, 'GN', 3372-3373, 'G', 3375, 'RLRHRRNAQNGPLSK
A;Cross-references: GB:L13458
A;Accession: B47648
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-546, 'P', 548-2198, 'D', 2290, 'NAR', 2294, 'L', 2296, 'WHATE', 2302-2303, 'V', 2305, '1
A;Title: 2516-2517, 'LOOG', 2522, 'IDG', 2526, 'S', 2528, 'SRGFHV', 2535, 'F', <RO2>
A;Cross-references: GB:L13458
A;Accession: C47648
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-546, 'P', 548-1128, 1290, 'DFARNSPS', 1299, 'NSS', 1303-1304, 'R', 'RHR', 1544-1545,
A;Cross-references: GB:L13458
C;Genetics;
A;Gene: CESP:ZC101.2e; CESP:ZC101.2a; CESP:ZC101.2c; CESP:ZC101.2b
A;Map position: 2
A;Introns: 32/1, 134/1, 225/1, 335/2, 450/3, 739/3, 830/3, 860/2, 1064/2, 1129/1, 1158/3

2; 2613/1; 2684/1; 2757/1; 2813/3; 2863/1; 2900/3; 3084/1; 3176/1; 3250/2
C;Superfamily: LR11 protein, laminin-type EGF-like homology; LDL receptor ligand-binding
C;Keywords: extracellular matrix
F;149-183/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;190-224/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;233-268/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;955-1002/Domain: laminin-type EGF-like homology <LEG1>
F;1011-1058/Domain: laminin-type EGF-like homology <LEG2>
Query Match 22.38; Score 349; DB 2; Length 3375;
Best Local Similarity 31.24; Pred. No. 4.5e-17;
Matches 81; Conservative 41; Mismatches 102; Indels 36; Gaps 6;
QY 30 GDLCEHSE--NPCQLRPFCLHGTCQCTR-----CLCLPGFSGPRCQCGSGHGAESDWH 82
DB 3129 GGVYSTQEPNNIC-ANSTCGMNGCVPRNTHYTCECKLYDGTCSLFK----- 3177
QY 83 LEGSGNDAPQCYCAYFHDGFLAPPGHVSRSPLPEVPTIELEVRTSTAGLLWQVE 142
DB 3178 -----PIEHAARFDGDAFIELSSDEFPHLTSEKDEIVAFKFKTEQQNGVLLWQ-Q 3227
QY 143 VGEAGQGKDFISLGLQGHVFRYQLSGSEARLVSEDPINDGEWHRVTALEGRGRSIQV 202
DB 3228 RPTVQOMEDYISVGVNGLHFSVELGGAHLISEERVDDGKEHSVRFERKREGQMRI 3287
QY 203 DCEELVSGRSPGNVAVNAKSVVIGAPDVATITGTFSSGITGCYXNLVHLHARPAP 262
DB 3288 DNYREVDGRSTGAILAMLNVGDNIFVGGVPDISKATGLFSNNFVGCIADELNGVK----- 3343
QY 263 PPQDLQLQRAQAQAGANTRPC 282
DB 3344 ----LDLMATAIDGNVQPC 3359
RESULT 4
AGRT
agrin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
C;Accession: JH0399; A38856
R;Rupp, F.; Payan, D.G.; Magill-Solc, C.; Cowan, D.M.; Scheller, R.H.
Neuron 6, 811-823, 1991
A;Title: Structure and expression of a rat agrin.
A;Reference number: JH0399; MUID:91222570; PMID:1851019
A;Accession: JH0399
A;Molecule type: mRNA
A;Residues: 1-1779;1799-1959 <RUP>
A;Cross-references: GB:M64780; NID:9202798; PIDN:AAA40703.1; PID:9202800
A;Experimental source: embryonic spinal cord
A;Note: it is uncertain whether Met-1, Met-18, or Met-24 is the initiator
R;Rupp, F.; Oezcelik, T.; Linial, M.; Peterson, K.; Francke, U.; Scheller, R.
J. Neurosci. 12, 3535-3544, 1992
A;Title: Structure and chromosomal localization of the mammalian agrin gene.
A;Reference number: A38856; MUID:92407628; PMID:1326608
A;Accession: A38856
A;Molecule type: mRNA
A;Residues: 1780-1798 <RU2>
A;Cross-references: GB:S44194
C;Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine re
C;Comment: 90% of rat embryonic transcripts encode the variant labeled below as form 3.1
C;Keywords: agrin; EGF homology; Katal proteinase inhibitor homology; laminin G repeat
C;Superfamily: agrin; EGF homology; Katal proteinase inhibitor homology; laminin G repeat
F;1-1959/Product: agrin, form 1 #status predicted <AG1>
F;1-1787, 1799-1959/Product: agrin, form 4 #status predicted <AG4>
F;1-1779, 1799-1959/Product: agrin, form 3 #status predicted <AG3>
F;1-1779, 1788-1959/Product: agrin, form 5 #status predicted <AG5>
F;1-1143, 1153-1959/Product: agrin, form 2 #status predicted <AG2>
F;22-50/Region: hydrophobic
F;88-137/Domain: Katal proteinase inhibitor homology <KPI1>
F;163-212/Domain: Katal proteinase inhibitor homology <KPI2>
F;236-284/Domain: Katal proteinase inhibitor homology <KPI3>
F;307-356/Domain: Katal proteinase inhibitor homology <KPI4>

F:381-429/Domain: Kazal proteinase inhibitor homology <KP15>
F:446-494/Domain: Kazal proteinase inhibitor homology <KP16>
F:511-559/Domain: Kazal proteinase inhibitor homology <KP17>
F:540-542/Region: motor neuron attachment (L-R-E) motif
F:596-645/Domain: Kazal proteinase inhibitor homology <KP18>
F:688-739/Domain: laminin-type EGF-like homology <LE1>
F:742-786/Domain: laminin-type EGF-like homology <LE2>
F:814-864/Domain: Kazal proteinase inhibitor homology <KP19>
F:869-992/Region: serine/threonine-rich
F:1094-1086/Region: motor neuron attachment (L-R-E) motif
F:1147-1215/Domain: EGF homology <EG1>
F:1287-1442/Domain: laminin G repeat homology <LG1>
F:1444-1476/Domain: EGF homology <EG2>
F:1483-1515/Domain: EGF homology <EG3>
F:1555-1706/Domain: laminin G repeat homology <LG2>
F:1713-1747/Domain: EGF homology <EG4>
F:1807-1959/Domain: laminin G repeat homology <LG3>
F:97-116, 105-137, 171-191, 180-212, 244-263, 252-284, 316-335, 324-356, 389-408, 397-429, 454-473
-1476, 1483-1494, 1488-1504, 1506-1515/Disulfide bonds: #status predicted
F:145, 672, 827, 957/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 21.3%; Score 333.5; DB 1; Length 1959;
Best Local Similarity 36.0%; Pred. NO. 3.4e-16;
Matches 94; Conservative 28; Mismatches 108; Indels 31; Gaps 10;

QY 1 CERQPCQHCATCMPAGEVEFCQCRDGFKGLDCEHENPCOLRPPCLHGTCQ-----GT 55
DB 1444 CUPNPHGALCQALEAGMFLCQCPGFRGPTCADEKSPCQ-PNFCGAACRVLSSGGA 1502

QY 56 RCLCLFGSGPRCQGGSHGIAESDWHLEGGGNDAPQGYGAYFHDDGFLAPFG-HVFSR 114
DB 1503 KCECLPLRSGTFCQT-----VLETAGSR-----PFLADFGFSLYELKGLHTFER 1547

QY 115 SLPEVPTIELEVRTSTAGLLWQGVVEGACQKDFISLGLQDGHVFRYQLGSGE 174
DB 1548 DLGE-KMALEWFLARGSGLLYNGQKTD--GKG-DFVSLNLRDGYLFKFKYDLGKGA 1603

QY 175 LVSEDPDNDGEWHRTALREGRSGIQVDGEELVSGRSPG----PNVAVNAKGSVYIGGA 230
DB 1604 IRSKEPIALGTWVRVFLERNRKGALQVGDGPRVLGSPKSRKVPDHTMLNKLPLEYIGGA 1663

QY 231 PDVATLT-GGRFSSGITGVCK 250
DB 1664 PDFSKLARGAAYSSGFSGVQ 1684

RESULT 5
agrin - electric ray (Discopyge ommata) (fragment)
C:Species: Discopyge ommata
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 17-Nov-2000
C:Accession: T43060
R:Smith, M.A.; Magill-Solc, C.; Rupp, F.; Yao, Y.M.M.; Schilling, J.W.; Snow, P.; McMahan, submitted to the EMBL Data Library, September 1992
A:Reference number: 22308
A:Accession: T43060
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1328 <SMI>
A:Cross-references: EMBL:L01423; NID:g213102; PID:g213103; PIDN:AAA49224.1
C:Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repeat
C:Keywords: glycoprotein; neuromuscular junction

Query Match 20.1%; Score 314.5; DB 2; Length 1328;
Best Local Similarity 31.4%; Pred. NO. 5.4e-15;
Matches 85; Conservative 44; Mismatches 105; Indels 37; Gaps 12;

QY 1 CERQPCQHCATCMPAGEVEFCQCRDGFKGLDCEHENPCOLRPPCLHGTCQ-----GT 55
DB 830 CSPNFCNKGKCHKMEKAEKFCESVGFSGPTCAKHNPDC-PNFCQSANCMWLPEGGS 888

QY 56 RCLCLPGSGPRCQGGSHGIAESDWHLEGGGNDAPQGYGAYFHDDGFLAPFG-HVFSR 114

DB 889 KCECPWGEGLCELR-----VSEAE-----QDQKAFIPETFNGLSYLWNGLHTF-- 933
QY 115 SLPEVPTIELEV--RTSTAGLLWQGVVEGACQKDFISLGLQDGHVFRYQLGSGE 172
DB 934 -VSDLLQKLSMEVIFLAKDPNGMFIYNGQKTD--GRG-DFVSLNLRDGYLFKFKYDLGKGA 989
QY 173 ARLVSEDPDNDGEWHRTALREGRSGIQVDGEELVSGRSP----GNVAVNAKGSVYIG 228
DB 990 AVRSKAPIPLNWNVTVERNRKGMLKINKDELVSSEPKSRKAPHTALNKEAFVVG 1049
QY 229 GAPD---VATLTGGRFSSGITGVCKNLVLHS 256
DB 1050 GAPDFNKPAAAG--IISGFTGAIOKLSLKS 1078

RESULT 6
AGCH
agrin precursor - chicken
C:Species: Gallus gallus (chicken)
C>Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
C:Accession: JH0591; A38857; E38857; I50692
R:Tsism, K.W.K.; Ruegg, M.A.; Escher, G.; Kroeger, S.; McMahan, U.J.
Neuron 8, 677-689, 1992
A:Title: cDNA that encodes active agrin.
A:Reference number: JH0591; MUID:92232297; PMID:1314620
A:Accession: JH0591
A:Molecule type: mRNA
A:Residues: 1-1955 <TSI>
A:Cross-references: GB:M94271; NID:g211120; PIDN:AAA48585.1; PID:g211121
A:Experimental source: brain
R:Ruegg, M.A.; Tsism, K.W.K.; Horton, S.E.; Kroeger, S.; Escher, G.; Gensch, E.M.; McMahan, Neuron 8, 691-699, 1992
A:Title: The agrin gene codes for a family of basal lamina proteins that differ in function
A:Reference number: A38857; MUID:92232298; PMID:1314621
A:Contents: alternative splicing
A:Accession: A38857
A:Molecule type: mRNA
A:Residues: 1132-1793; 1795-1955 <RU2>
A:Cross-references: GB:M97371
A:Accession: B38857
A:Molecule type: mRNA
A:Residues: 1221-1647; 1652-1783; 1794-1955 <RU3>
A:Cross-references: GB:M97372
A:Note: translation of the nucleotide sequence is not complete
R:Thomas, W.S.; O'Dowd, D.K.; Smith, M.A.
Dev. Biol. 158, 523-535, 1993
A:Title: Developmental expression and alternative splicing of chick agrin RNA.
A:Reference number: I50692; MUID:93345745; PMID:8393816
A:Accession: I50692
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: SHLSNEIPA, 1784-1795 <THO>
A:Cross-references: EMBL:U07271; NID:g459665; PIDN:AAAL6788.1; PID:g459666
C:Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine receptor subunits. Alternative splicing produces two inactive proteins: agrin-related protein 1 and agrin-related protein 2.
C:Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repeat
C:Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction

F:1-38/Domain: signal sequence #status predicted <SIG>
F:139-1955/Product: agrin #status predicted <MAN>
F:139-1955/Product: agrin-related protein 1 #status predicted <AG1>
F:139-1955/Product: agrin-related protein 2 #status predicted <AG2>
F:139-1647, 1652-1783, 1794-1955/Product: agrin-related protein 1 #status predicted <AG3>
F:177-126/Domain: Kazal proteinase inhibitor homology <KPI1>
F:152-201/Domain: Kazal proteinase inhibitor homology <KPI2>
F:225-273/Domain: Kazal proteinase inhibitor homology <KPI3>
F:295-344/Domain: Kazal proteinase inhibitor homology <KPI4>
F:370-418/Domain: Kazal proteinase inhibitor homology <KPI5>
F:433-483/Domain: Kazal proteinase inhibitor homology <KPI6>
F:500-548/Domain: Kazal proteinase inhibitor homology <KPI7>
F:584-633/Domain: Kazal proteinase inhibitor homology <KPI8>
F:675-726/Domain: laminin-type EGF-like homology <LE1>
F:729-773/Domain: laminin-type EGF-like homology <LE2>
F:801-851/Domain: Kazal proteinase inhibitor homology <KPI9>

F:855-995/Region: serine/threonine-rich
F:1150-1219/Region: serine/threonine-rich
F:1233-1264/Domain: EGF homology <EG1>
F:1294-1448/Domain: laminin G repeat homology <LG1>
F:1429-1433/Region: motor neuron attachment (L-R-E) motif
F:1450-1482/Domain: EGF homology <EG2>
F:1489-1521/Domain: EGF homology <EG3>
F:1560-1711/Domain: laminin G repeat homology <EG4>
F:1718-1751/Domain: EGF homology <EG5>
F:1803-1955/Domain: laminin G repeat homology <LG2>
F:86-105, 94-126, 150-180, 233-252, 241-273, 304-323, 312-344, 378-397, 386-418, 443-462,
F:1489-1500, 1494-1510, 1512-1521/Disulfide bonds: #status predicted
F:390,659,764,814/Binding site: carbohydrate (Asn) #status predicted

Query Match 19.3%; Score 301.5; DB 1; Length 1955;
Best Local Similarity 27.4%; Pred. No. 7.1e-14;
Matches 87; Conservative 38; Mismatches 108; Indels 85; Gaps 13;

QY 1 CERPOCHGATCMPAGEVEFOCLCRDGFKGLCHEENPCQLREPCLLHGTC-----QGT 55
DB 1450 CHNPHCHGASCHVKEAMFHCHCLHSYTGTCADERNPTD-PTPCHISATCLVLPEGGA 1508
QY 56 RCLCLPFGSPRCQCGSHGIAESDWHLEGGSGNDAPGQYGAYFHDGFLAFPGHVFSRS 115
DB 1509 MCACPMRGEGFCER-----VTEQD-----HT 1530
QY 116 LPETPE-----TIELEVRTSTASGLLLWQGEVGEAGQKDFI 153
DB 1531 MPFLPEFNGFSYELNGLOTFLTCRQMSMEVFLAKSPGMIFYNQKTD--GKG-DFV 1587
QY 154 SLGQDGLHVPYRLGSGEARLVSEDPINDGEHVRVLTALREGRGSIQVGDGELVGRSP 213
DB 1568 SLALHDGYLEYRDLGKGAVALRSKEFVPLNTWISLVLLERSGKGVNRINNGERVNGESP 1647
QY 214 G-----PNVAVNAKSVYIGGAPDVATLT--GQFSSGITGCVKNLVLSRAPGPPQPLD 268
DB 1648 KSRKVPFAPLNLKPFYVGGAPDFSKLARAATISFYGAVQIRISI-----KGVP----LLK 1700
QY 269 LQH-RAQAGANT---RPC 282
DB 1701 EQHRSAYEISTFRAHPC 1718

RESULT 7
T42218
slit-1 protein homolog - rat
N:Alternate names: MEFG4 protein
C:Species: Rattus norvegicus (Norway rat)
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 16-Aug-2002
C:Accession: T42218
R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A:Reference number: Z14126; MUID:98360089; PMID:9693030
A:Accession: T42218
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1531 <NAK>
A:Cross-references: EMBL:AB011530; NID:G3449289; PIDN:BA32460.1; PID:G3449290
A:Experimental source: strain Sprague-Dawley; brain
C:Genetics:
A:Gene: MEFG4
C:Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein

Query Match 18.1%; Score 283.5; DB 2; Length 1531;
Best Local Similarity 27.9%; Pred. No. 1.1e-12;
Matches 81; Conservative 48; Mismatches 92; Indels 69; Gaps 14;

QY 1 CERPOCHGATCMPAGEVEFOCLCRDGFKGLCHEENPCQLREPCLLHGTC-----QGT 55
DB 1087 KDHQCCQNGAQCQVDEIN-SIACLCAGISGQGLCEIPAPRNSCEGTE-CQNGANCVDGS 1144
QY 56 R--CLCLPFGSPRCQCGSHGIAESDWHLEGGSGNDAPGQYGAYFHDGFLAFPGHVFS 113

DB 1145 RPYQCCLFPGFGPECEKLLSNFVDRDITYLQ-----FTD----- 1178
QY 114 RSLPEVPE-TIELEVRTSTASGLLLWQGEVGEAGQKDFISLGLQDGLHVRVOLGS-G 171
DB 1179 --LQNWPRANITLQVSTAEDNGILLNG-----DNDHIAVELYQGHVRSYDPGSGP 1228
QY 172 EARLYSEDPINDGEHVRVLTALREGRGSIQVGDGELVSGRSPGNVAVNAKGSVYIGGAP 231
DB 1229 SSAIYSAETINDQGHVTVLFTFQMVNLSDGGSPTMDFNGKHVTLNSRAPLYVGMP 1288
QY 232 -DVAT-----LTGGRFSSGITGCVKNLVLS-----ARPGAPP 263
DB 1289 VDVNSAARLWQILNGTSFH---GCIRNLVNNELQDFTKTQMKPGVVP 1334

RESULT 8
T20721
hypothetical protein F25F2.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 19-May-2000
C:Accession: T20721; T21343; T21342
R:Ainscough, R.
submitted to the EMBL Data Library, August 1994
A:Reference number: Z19314
A:Accession: T20721
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4307 <WIL>
A:Cross-references: EMBL:Z35662; PIDN:CAA84721.1; GSPDB:GN00021; CESP:F25F2.2
A:Experimental source: clone F10G11
R:Ainscough, R.
submitted to the EMBL Data Library, August 1994
A:Reference number: Z19410
A:Accession: T21343
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4307 <W12>
A:Cross-references: EMBL:Z35599; PIDN:CAA84661.1; GSPDB:GN00021; CESP:F25F2.2
A:Experimental source: clone F25F2
R:Sulston, J.
submitted to the EMBL Data Library, June 1994
A:Reference number: Z19806
A:Accession: T23842
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4307 <W13>
A:Cross-references: EMBL:Z34802; PIDN:CAA84339.1; GSPDB:GN00021; CESP:F25F2.2
A:Experimental source: clone M88
C:Genetics:
A:Gene: CESP:F25F2.2
A:Map position: 3
A:Introns: 79/3; 167/3; 251/3; 328/3; 587/3; 625/3; 737/3; 1006/1; 1190/2; 1494/2; 1770/3;
C:Superfamily: rat MEFG1 protein; cadherin repeat homology; EGF homology; laminin G repeat
F:3761-3900/Domain: laminin G repeat homology <LGR>

Query Match 16.9%; Score 264; DB 2; Length 4307;
Best Local Similarity 26.4%; Pred. No. 8.6e-11;
Matches 81; Conservative 45; Mismatches 99; Indels 82; Gaps 13;

QY 21 QCLCRDGFKGLCHEENPC-----QLRPFCLHGTCQCTRCCLCLPFGSPRCQ--- 69
DB 3670 QCLCPDGFQGRKEVETNQCAKSPCQWQLCIPSVHNSTYE---CVCPLMGEGDKCVSPS 3726
QY 70 -QSGSHGIAESDWHLEGGSGNDAPGQYGAYFHDGFLAFPGHVFSRSLPEVPE---IEL 125
DB 3727 CQNDGKCLREELSVCGDG-----YFEISLSNELETRMELEI 3763
QY 126 EVRTSTASGLLLWQGEVGEAGQK-DFISLGLQDGLHVRVOLGSGEARLVSEDPINDG 184
DB 3764 ELAKTTTNGIIMW-----SRGKRDFHMLRVNGTPEYHWNAGTGTGIVTSKTSVWDG 3815
QY 185 EHWRTVLTALREGRGSIQVGDGELVSGRSPGNVAVNA---KGSVYIGGAPDVATLTGGRF 241

Db 347 VSAELYDGRVXLYVYIGNFPASHMYSSVKVNDGLPHRISIRTSERKCFQLQIDKNPQVIVE 406
QY 210 GRSPGNVAVNAKGSVYIGGAP-----DVATLTGRRPSSGITGCVKVLVHSARPQAPP 264
Db 407 NSGKSDQITGKEMLYIGGLPIEKSDQAKRRFHVKNSESLKGCISITINEV-----459
QY 265 QPLDLQHPAQAGANT 279
Db 460 -PINLQ-QALENVNT 472
RESULT 11
D89711
protein F40E10.4 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001
C:Accession: D89711
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A:Reference number: A75000; MUID:99069613; PMID:1981916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: D89711
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-601 <STO>
A:Cross-references: GB:chr_X; PIDN:CAA93668.1; PID:g3877014; GSPDB:GN00028; CESP:F40E10.
C:Genetics:
A:Gene: F40E10.4
A:Map position: X

Query Match 15.9%; Score 249; DB 2; Length 601;
Best Local Similarity 26.7%; Pred. No. 1.3e-10;
Matches 84; Conservative 48; Mismatches 115; Indels 68; Gaps 17;
QY 5 PCQCHGATCMA-GRYEFQCLCRDGFKGLDCHENPQCLREPLHGGTC-----QGTRCLC 59
Db 186 PCENNGKCIPIKNGSY--SCMCPGFTGNCTETIDCKNVE--CONGGSCVDGILSYDCLC 242
QY 60 LPFGSG-----PRCQ-----GSGHGIA---ESDWHL---EGSGNDAPQ 94
Db 243 RFGVAGYCEIPPMWMEYQKTDACQSQAGCBVASQNSDFTCKCHEGSPGSCDRQ 302
QY 95 YGAYPHDGG-ELAPGPHVFSRSLPEVE-TIELEVRTSTAGLLMQGVVEGAGQKDF 152
Db 303 MSVGFKNFGAYLAD-----FLASDGTITMTLTFTTSKIGILYYGDD-----HF 346
QY 153 ISLGLQDGLHLYRYQLSGSEA-RLVSEDPINDGSHWRVTALREGRSGSIQVDGE--ELVS 209
Db 347 VSAELYDGRVXLYVYIGNFPASHMYSSVKVNDGLPHRISIRTSERKCFQLQIDKNPQVIVE 406
QY 210 GRSPGNVAVNAKGSVYIGGAP-----DVATLTGRRPSSGITGCVKVLVHSARPQAPP 264
Db 407 NSGKSDQITGKEMLYIGGLPIEKSDQAKRRFHVKNSESLKGCISITINEV-----459
QY 265 QPLDLQHPAQAGANT 279
Db 460 -PINLQ-QALENVNT 472

RESULT 12
T13953
MEGF5 protein - rat
N:Alternate names: slit protein homolog
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 16-Aug-2002
C:Accession: T13953
R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genetics 51, 27-34, 1998
A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A:Reference number: Z14126; MUID:99360089; PMID:9693030
A:Accession: T13953

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1523 <NAK>
A:Cross-references: EMBL:AB011531; NID:g3449291; PIDN:BAA32461.1; PID:g3449292
C:Genetics:
A:Gene: MEGF5
C:Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein r
Query Match 15.7%; Score 246; DB 2; Length 1523;
Best Local Similarity 27.1%; Pred. No. 5.8e-10;
Matches 79; Conservative 40; Mismatches 106; Indels 66; Gaps 13;
QY 1 CEROPCOHGATCMPAGEYEFQCLCRDGFKGLDCEH-----EENPCOLREPLHGGTC- 52
Db 1078 CVAHKCRHGAQCDAVN-GYTICPCQGFSGLFCEHPMPVLLQTSPCDQVE--CONGAQCI 1135
QY 53 ---QGTRCLCLPQSGPRCCQSGSGHGAESDWHLESGGNDAPQYQAYFHDGFLAFPG 109
Db 1136 VVQPEPTCRCPGAGPRCKELITVNFVGKDSVELASAKVRP-----1178
QY 110 HVFSRSLPEVPEITIELEVRTSTAGLLMQGVVEGAGQKDFISLGLQDGLHLYRY-OL 168
Db 1179 -----QANISLQVATDKONGILLYKG-----DNDPLALELYQGHVRLYDLSL 1220
QY 169 GSGEARLVSEDPINDGSHWRVTALREGRSGSIQVD--GEELVSGR--SPGPNVAVNAKGSVY 226
Db 1221 SSPPTTVYSVETVNDGQFVSHVLMNLQTLNVLDKAPKSLQKQKQPAVGINS--PLY 1278
QY 227 IGGAPDVATITG-----GRFSSGITGCVKVLVHSARPQ--APPOPL 267
Db 1279 LGGTP---TSTGLSALRQAGDRFLGGFHGCTHEVRINNELQDFKALPPOS 1326

RESULT 13
T42626
secreted leucine-rich repeat-containing protein SLIT2 - mouse (fragment)
N:Alternate names: neurogenic extracellular slit protein
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 16-Aug-2002
C:Accession: T42626
R:Holmes, G.P.; Negus, K.; Burridge, L.; Raman, S.; Algar, E.; Yamada, T.; Little, M.H.
Mech. Dev. 79, 57-72, 1998
A:Title: Distinct but overlapping expression patterns of two vertebrate slit homologs im
A:Reference number: 222177; MUID:99279238; PMID:10349621
A:Accession: T42626
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1025 <HOL>
A:Cross-references: EMBL:AF074960; NID:g4151258; PID:g4151259; PIDN:AAD04345.1
C:Genetics:
A:Gene: Slit2
C:Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein r
Query Match 15.5%; Score 243.5; DB 2; Length 1025;
Best Local Similarity 24.8%; Pred. No. 5.8e-10;
Matches 92; Conservative 45; Mismatches 103; Indels 131; Gaps 18;
QY 1 CEROPCOHGATC-MPAGEYE-FOCLCRDGFKGLDCE-----HEENPC-----40
Db 457 CISNPKCHGCTCHKEGAGFWCTCADOFESENCVENLDDCEDNCCNNSCTVDGINNY 516
QY 41 -----QLREPLHGGTC-----QGTRCLCLPQSGP-----66
Db 517 TCLCPPEYTGELCEEKLDFOADLNPCQDHSKILTPKGFKCDCTPGYIGEHCDIDFDC 576
QY 67 ---RCQCGSGHGAESDWHL---EGSGG-----NDA-----91
Db 577 QDNKCKNGAHCCTDAVNGYTCVCPGEGYSGLFCEFSPPVLPRTSPCDNFCQNGAQCIIRI 636
QY 92 -----PQYGA-----YFHDGFLAFPGHVFPSRSLPEVPEP-TIELEVRTSTAS 133
Db 637 NEPICQCLPGYLGEKCEKLVSNFVNKESYLQTP-----SAKVRPQTNITLIQIATDEDS 690

QY 134 GLLWQGVGEAGQKQDFTSLGQDGHVFRYQLGSGEARLV-SBDPINDGEMERVTAL 192
 Db 691 GILLYKG-----DKDHLAVELRYGRVRAASVDTGSHPSAASVYVETINDGNFHVILL 742
 QY 193 REGRRGSIQVDG--EELVSGRSQPNVAVNAKGSVYIGAP--DVATL--TGGRFSSGI 245
 Db 743 TLDSSLSLSDVGGSPKVTNLS--KQSTLNFDSPFLVGGPGKNVAVSLRQAPGQNGTSF 800
 QY 246 TGCYNLVLHS 256
 Db 801 HGCTIRNLYINS 811

RESULT 14
 MMHUMH
 laminin alpha-2 chain - human (fragment)
 A/Alternate names: laminin M chain; merosin heavy chain
 C/Species: Homo sapiens (man)
 C/Date: 30-Sep-1991 #sequence_revision 21-Aug-1998 #text_change 10-Dec-1999
 C/Accession: PX0082; A35899; A38970; S14461
 R/Hori, H.; Kanamori, T.; Mizuta, T.; Yamaguchi, N.; Liu, Y.; Nagai, Y.
 J. Biochem. 116, 1212-1219, 1994
 A/Title: Human laminin M chain: Epitope analysis of its monoclonal antibodies by immunoblotting
 A/Reference number: PX0082; MUID:95221315; PMID:7535762
 A/Accession: PX0082
 A/Molecule type: mRNA
 A/Residues: 1-1751 <HOR>
 A/Experimental source: placenta
 R/Ehrig, K.; Leivo, I.; Argaves, W.S.; Ruoslahti, E.; Engvall, E.
 Proc. Natl. Acad. Sci. U.S.A. 87, 3264-3268, 1990
 A/Title: Merosin, a tissue-specific basement membrane protein, is a laminin-like protein
 A/Reference number: A35899, MUID:90238994; PMID:2185464
 A/Accession: A35899
 A/Status: nucleic acid sequence not shown
 A/Molecule type: mRNA
 A/Residues: 'V', 623-1751 <EHR1>
 A/Cross-references: EMBL:M59832
 A/Accession: A38970
 A/Molecule type: protein
 A/Residues: 1368-1384; 1389-1406; 1593-1607 <EHR2>
 A/Note: the sequence from fig. 1 is inconsistent with that from fig. 2 in lacking 1599-1607
 R/Ehrig, K.; Leivo, I.; Argaves, S.W.; Ruoslahti, E.; Engvall, E.
 submitted to the EMBL Data Library, December 1990
 A/Description: The tissue-specific basement membrane protein merosin is a laminin-like protein
 A/Reference number: S14461
 A/Accession: S14461
 A/Molecule type: mRNA
 A/Residues: 'V', 623-1264, 'R', 1266-1751 <LE1>
 A/Cross-references: EMBL:M59832; NID:G187520; PIDN:AA63215.1; PID:G187521
 C/Comment: This protein is a prominent component of the basement membrane that mediates cell-matrix interactions
 C/Genetics:
 A/Gene: GDB:LAWA2; LAMM
 A/Cross-references: GDB:132362; OMIM:156225
 A/Map position: 6Q22-6Q23
 C/Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
 C/Function:
 A/Description: interact with cells and with other basement membrane proteins to promote cell-matrix interactions
 C/Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like repeat homology
 C/Keywords: basement membrane; calcium binding; coiled coil; extracellular matrix; Glyco
 F:21-58/Domain: laminin-type EGF-like homology #status atypical <LE01>
 F:61-107/Domain: laminin-type EGF-like homology <LE02>
 F:110-165/Domain: laminin-type EGF-like homology <LE03>
 F:168-212/Domain: laminin-type EGF-like homology <LE04>
 F:527-567, 1071-1300/Region: 3DM and 2D9 binding
 F:811-972/Domain: laminin G repeat homology <LG1>
 F:1005-1165/Domain: laminin G repeat homology <LG2>
 F:1131-1254/Domain: laminin G repeat homology <LG3>
 F:1430-1578/Domain: laminin G repeat homology <LG4>
 F:1605-1751/Domain: laminin G repeat homology <LG5>
 F:120,238,255,341,451,542,557,561,658,666,767,881,1001,1076,1119,1192,1199,1289,1509

Query Match 14.3%; Score 223.5; DB 1; Length 1751;
 Best Local Similarity 36.4%; Pred. No. 2.9e-08;

Matches 59; Conservative 23; Mismatches 67; Indels 13; Gaps 6;
 QY 122 TIEVTRTASGILLWQGVGEAGQKQDFTSLGQDGHVFRYQLGSGEARLVSEDP 181
 Db 1429 TIEVTRTEASGLLFYM-ARINHA---DPATVQLRNLGLFVSYDLGSGDTHMPTKI 1483
 QY 182 NDEWHRVATLREGRRGSIQVDGSELVSGRSQPNVAVNAKGSVYIGAP-DVATLTG 238
 Db 1484 NDCQWHKIKIMRSKQEGILYVDG---ASNRTISPCKADILDVVGMLYVGGLPFYNTTTRI 1540
 QY 239 GRFSSGITGCVKNLVLHSARPQAPPPQPLDLQHRQAQGANTR 280
 Db 1541 GPVTYSIDGCVN--LHMAEPADLEQPTSSSFHVGTCPANAQ 1580

RESULT 15
 A48216
 neurexin III-alpha secreted type I precursor - rat
 C/Species: Rattus norvegicus (Norway rat)
 C/Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 09-Dec-2002
 C/Accession: A48216; B48216
 R/Ushkaryov, Y.A.; Suedhof, T.C.
 Proc. Natl. Acad. Sci. U.S.A. 90, 6410-6414, 1993
 A/Title: Neurexin IIIalpha: extensive alternative splicing generates membrane-bound and soluble isoforms
 A/Reference number: A48216; MUID:93342001; PMID:8341647
 A/Accession: A48216
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-1438 <USH>
 A/Cross-references: GB:U14851
 A/Accession: B48216
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-1368, 1372-1438 <US2>
 A/Cross-references: GB:U14851
 C/Genetics:
 A/Introns: 1372/1
 C/Superfamily: neurexin; EGF homology
 C/Keywords: alternative splicing; brain; cell surface component; duplication; extracellular matrix
 F:1-27/Domain: signal sequence #status predicted <SIG>
 F:202-234/Domain: EGF homology <EGF>
 F:651-683/Domain: EGF homology <EGF1>

Query Match 13.9%; Score 217; DB 2; Length 1438;
 Best Local Similarity 26.4%; Pred. No. 6.9e-08;
 Matches 67; Conservative 35; Mismatches 114; Indels 38; Gaps 8;
 QY 27 GFKGLDCE-----HEENPQLEPCPLHGTC-----QGTRCLC-LPGF 63
 Db 171 GFKGLMDLKYNSPRLGSSVQLAEQPCGER-PCENGGICPLLDGCHTCDCTTGY 229
 QY 64 SGPRCQGGSGHGAESDMHLEGGSGNDAPQGVAGYFHDGDFLAPGHVFSRLSEVPETI 123
 Db 230 GGTLCSEDSVQSGPLSHLMMSQGRSKAREENAVTFRGSEYLCV--DLSONPIQSSDEI 287
 QY 124 ELVTRTASGILLWQGVGEAGQKQDFTSLGQDGHVFRYQLGSGEARLVSEDP 181
 Db 288 TLFKTYWRNGILL-----HTGSAVYVNLAKDGAVALINLGSAGFAIVE-PVNG 339
 QY 182 --NDGEWHRVATLREGRRGSIQVDGSELVSGRSQPNVAVNAKGSVYIGAPDVATLTG 239
 Db 340 KFDNNAHDVKTTRNLQRTVISVDGILTTTGTQSDYTMGSDDDFFYVGGSPSTADLP 399
 QY 240 RFSSGITGCVKNLVL 253
 Db 400 PVSNFPGCLKEV 413

Search completed: March 9, 2004, 17:22:35
 Job time : 10.7114 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:19:08 ; Search time 20.844 Seconds
(without alignments)
2876.963 Million cell updates/sec

Title: US-10-006-011a-9

Perfect score: 1566

Sequence: 1 CERQPCQHGATCMRAGEYEF.....QPLDLQHRAGANTRCPS 284

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	357.5	22.8	1931	15	US-10-094-886-180
2	343	21.9	238	9	US-09-764-853-758
3	343	21.9	238	9	US-09-764-898-252
4	343	21.9	238	10	US-09-764-881-158
5	343	21.9	238	14	US-10-073-865-121
6	343	21.9	238	15	US-10-242-747-158
7	343	21.9	375	9	US-09-978-249-10
8	343	21.9	432	9	US-09-764-853-541
9	343	21.9	432	9	US-09-764-898-179
10	343	21.9	432	10	US-09-764-881-102
11	343	21.9	432	14	US-10-073-865-78
12	343	21.9	432	15	US-10-242-747-102
13	343	21.9	463	15	US-10-104-047-3058
14	343	21.9	775	15	US-10-108-260A-4433
15	340.5	21.7	406	15	US-10-264-049-2933

Sequence 34, Appl	US-10-016-283-34	1940	13	333.5	21.3	1940	13	US-10-016-283-34	Sequence 34, Appl
Sequence 29, Appl	US-10-016-283-29	294	13	330	21.1	294	13	US-10-016-283-29	Sequence 29, Appl
Sequence 28, Appl	US-10-016-283-28	338	13	330	21.1	338	13	US-10-016-283-28	Sequence 28, Appl
Sequence 27, Appl	US-10-016-283-27	390	13	330	21.1	390	13	US-10-016-283-27	Sequence 27, Appl
Sequence 26, Appl	US-10-016-283-26	440	13	330	21.1	440	13	US-10-016-283-26	Sequence 26, Appl
Sequence 25, Appl	US-10-016-283-25	456	13	330	21.1	456	13	US-10-016-283-25	Sequence 25, Appl
Sequence 36, Appl	US-10-016-283-36	492	13	330	21.1	492	13	US-10-016-283-36	Sequence 36, Appl
Sequence 503, Appl	US-09-866-050A-503	819	10	330	21.1	819	10	US-09-866-050A-503	Sequence 503, Appl
Sequence 30, Appl	US-10-016-283-30	256	13	326	20.8	256	13	US-10-016-283-30	Sequence 30, Appl
Sequence 30, Appl	US-09-970-944-30	1534	11	292	18.6	1534	11	US-09-970-944-30	Sequence 30, Appl
Sequence 432, Appl	US-09-764-870-432	210	9	290.5	18.6	210	9	US-09-764-870-432	Sequence 432, Appl
Sequence 432, Appl	US-10-125-540-432	27	14	290.5	18.6	27	14	US-10-125-540-432	Sequence 432, Appl
Sequence 31, Appl	US-09-970-944-31	1531	11	283.5	18.1	1531	11	US-09-970-944-31	Sequence 31, Appl
Sequence 28, Appl	US-09-970-944-28	1531	11	280.5	17.9	1531	11	US-09-970-944-28	Sequence 28, Appl
Sequence 338, Appl	US-10-104-047-338	152	15	278.5	17.8	152	15	US-10-104-047-338	Sequence 338, Appl
Sequence 4, Appl	US-10-174-677-4	5147	14	269	17.2	5147	14	US-10-174-677-4	Sequence 4, Appl
Sequence 5698, Appl	US-10-369-493-5698	4307	15	264	16.9	4307	15	US-10-369-493-5698	Sequence 5698, Appl
Sequence 5699, Appl	US-10-369-493-5699	4307	15	264	16.9	4307	15	US-10-369-493-5699	Sequence 5699, Appl
Sequence 5700, Appl	US-10-369-493-5700	4307	15	264	16.9	4307	15	US-10-369-493-5700	Sequence 5700, Appl
Sequence 31, Appl	US-10-016-283-31	216	13	261	16.7	216	13	US-10-016-283-31	Sequence 31, Appl
Sequence 2, Appl	US-10-289-776-2	1525	14	253.5	16.2	1525	14	US-10-289-776-2	Sequence 2, Appl
Sequence 29, Appl	US-09-766-511B-29	1529	10	253.5	16.2	1529	10	US-09-766-511B-29	Sequence 29, Appl
Sequence 67, Appl	US-10-189-123-67	1529	14	253.5	16.2	1529	14	US-10-189-123-67	Sequence 67, Appl
Sequence 537, Appl	US-10-188-495-67	1529	14	253.5	16.2	1529	14	US-10-188-495-67	Sequence 537, Appl
Sequence 537, Appl	US-09-764-870-537	194	9	252	16.1	194	9	US-09-764-870-537	Sequence 537, Appl
Sequence 6999, Appl	US-10-125-540-537	194	14	252	16.1	194	14	US-10-125-540-537	Sequence 6999, Appl
Sequence 9, Appl	US-10-369-493-6999	601	15	249	15.9	601	15	US-10-369-493-6999	Sequence 9, Appl
Sequence 396, Appl	US-10-289-776-9	735	14	249	15.9	735	14	US-10-289-776-9	Sequence 396, Appl
Sequence 32, Appl	US-09-866-050A-396	1529	10	245.5	15.7	1529	10	US-09-866-050A-396	Sequence 32, Appl
	US-09-970-944-32	1512	11	244.5	15.6	1512	11	US-09-970-944-32	

ALIGNMENTS

RESULT 1
US-10-094-886-180
Sequence 180, Application US/10094886
Publication No. US20040002120A1

GENERAL INFORMATION:
APPLICANT: Kekuda, Ramesh

APPLICANT: Tchernev, Velizar T.
APPLICANT: Liu, Xiaohong

APPLICANT: Spytek, Kimberly A.
APPLICANT: Patturajan, Meeta

APPLICANT: Burgess, Catherine
APPLICANT: Vernet, Corine A.

APPLICANT: Li, Li
APPLICANT: Gorman, Linda

APPLICANT: Malyankar, Uriel M.
APPLICANT: Boldog, Ferenc

APPLICANT: Guo, Xiaojia
APPLICANT: Shenoy, Suresh

APPLICANT: Padigaru, Muralidhara
APPLICANT: Taupier, Raymond J., Jr.

APPLICANT: Miller, Charles
APPLICANT: Casman, Stacie

APPLICANT: Pena, Carol
APPLICANT: Gangolli, Esha

APPLICANT: Gusev, Vladimir
APPLICANT: Smithson, Glenda

APPLICANT: Zerhusen, Bryan
APPLICANT: Gerlach, Valerie

APPLICANT: Pochart, Pascal
APPLICANT: Fernandes, Elma

APPLICANT: Shimkets, Richard
APPLICANT: Rastelli, Luca

APPLICANT: Spaderna, Steven
APPLICANT: LaRoche, William

APPLICANT: Zhong, Mei
TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD

FILE REFERENCE: 21402-290 B
CURRENT APPLICATION NUMBER: US/10/094, 886

```
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/274,322
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,182
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/288,052
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/318,510
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/314,018
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/296,693
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/313,626
; PRIOR FILING DATE: 2001-08-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: Patent in 2.1
; SEQ ID NO 180
; LENGTH: 1931
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-886-180

Query Match      22.8%; Score 357.5; DB 15; Length 1931;
Best Local Similarity 35.4%; Pred. No. 2.3e-22;
Matches 91; Conservative 29; Mismatches 110; Indels 27; Gaps 8;

Qy 1 CERPCOHGATCPAGEYEQCLCRDGFKXGDLCEHEENPCQLREPCPLHGTCQ-----GT 55
Db 1439 CLNPFCHGAPCNLEAGRHFCQCPGPRVGPCTCADEKSPCQ-PNCHGAAPCRVLPEGA 1497

Qy 56 RCLCLPFCSPRQCGSGHGIASDWHLESGNGNDARGQYGFYHDDGFLAPPG-HVFSR 114
Db 1498 QCECLRGRTFCQTAS-----GQDGGPFLADFNFGHLELRGLHTFAR 1542

Qy 115 SLPEVPETIEVTRTSTAGLLMQGVEVGEAGQGXDFISLGLQDGHLYFRYQLGSGEAR 174
Db 1543 DLGE-KMALEWFLARGPSGLLYNGQKTD--GKG-DFVSLALRRRLFRYDLGKGA 1598

Qy 175 LVSEDPINDGEHVRVTRALRGSGSIQVDGEELVSGSPGNVAVNAKGSVYVIGGAPDA 234
Db 1599 IRSREPTVLGAWTRVSLERNRKGALRVGDPFRVLGSPVPHVTLNKLPELVGGAPDFS 1658

Qy 235 TLT-GRFSSGITGVCK 250
Db 1659 KLARAAVSSGFDGAIQ 1675

RESULT 2
US-09-764-853-758
; Sequence 758, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: R206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 758
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

```
US-09-764-898-252
; Sequence 252, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: R201
; CURRENT APPLICATION NUMBER: US/09/764,898
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 311
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 252
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-898-252

Query Match      21.9%; Score 343; DB 9; Length 238;
Best Local Similarity 33.5%; Pred. No. 3.1e-22;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

Qy 34 EHEENPCQLREPCPLHGTC-----QGTCLCLPFGSPRQCGSGHGIASDWHLESGNGN 89
Db 11 ENAAHPC-VRAPCAHGGSCRPKEGYDCDCPLGPEGLHCOKAIEAL-----56

Qy 90 DAP-----GOYGAYFHDDGFLAPPGHVFERSLPEVPEETIEVTRTSTAGLLMQGVEVGE 145
Db 57 EIPQFGRSYLTYPDILKRVSG--SR-----NVFMFKTTAKDGLLLWRG--DSP 105

Qy 146 AGGQKDFISLGLQDGHLYFRYQLGSGEARLYSEDPINDGEHVRVTRALRGSGSIQVDGE 205
Db 106 MRPNDFISLGLRDGALVFSYNLGSVASIMVNGSFNDGRHVRVKAQVDSGSGKITVDDY 165

Qy 206 ELVSGRSPGNVAVNAKGSVYVIGGAPDAVATLTGGRFSSGITGVCKNLVLSHARFGAPPPQ 265
Db 166 GARTGKSPGMFQNLINGALVGGMKEIALHTNRQYMRGLVGCISHFTLST-----DY 218

Qy 266 PLDLQHRQAQAGANTRPC 282
Db 219 HISLVEDAVDGKNINTC 235
```

```
RESULT 3
US-09-764-898-252
; Sequence 252, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: R201
; CURRENT APPLICATION NUMBER: US/09/764,898
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 311
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 252
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-898-252

Query Match      21.9%; Score 343; DB 9; Length 238;
Best Local Similarity 33.5%; Pred. No. 3.1e-22;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

Qy 34 EHEENPCQLREPCPLHGTC-----QGTCLCLPFGSPRQCGSGHGIASDWHLESGNGN 89
Db 11 ENAAHPC-VRAPCAHGGSCRPKEGYDCDCPLGPEGLHCOKAIEAL-----56

Qy 90 DAP-----GOYGAYFHDDGFLAPPGHVFERSLPEVPEETIEVTRTSTAGLLMQGVEVGE 145
Db 57 EIPQFGRSYLTYPDILKRVSG--SR-----NVFMFKTTAKDGLLLWRG--DSP 105

Qy 146 AGGQKDFISLGLQDGHLYFRYQLGSGEARLYSEDPINDGEHVRVTRALRGSGSIQVDGE 205
Db 106 MRPNDFISLGLRDGALVFSYNLGSVASIMVNGSFNDGRHVRVKAQVDSGSGKITVDDY 165

Qy 206 ELVSGRSPGNVAVNAKGSVYVIGGAPDAVATLTGGRFSSGITGVCKNLVLSHARFGAPPPQ 265
Db 166 GARTGKSPGMFQNLINGALVGGMKEIALHTNRQYMRGLVGCISHFTLST-----DY 218
```

QY 266 PDLQHRQAQAGANTRPC 282
Db 219 HISLVEDAVDGKNINTC 235

RESULT 4

US-09-764-881-158

; Sequence 158, Application US/09764881
; Publication No. US20030125246A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ07
; CURRENT APPLICATION NUMBER: US/09/764,881
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 192
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-881-158

Query Match 21.9%; Score 343; DB 10; Length 238;
Best Local Similarity 33.5%; Pred. No. 3.1e-22;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
QY 34 EHENPCOLREPLHGTC-----QTRCLCLPFGSGRCQSGHGHIAESDWHLESGGN 89
Db 11 ENAAHPC-VRAPCAHGSCSRKEGYDCDCPLGPEGLHCOKAITEAI-----56
QY 90 DAP-----GOYGAYFDDGFLAFPGHVFERSLPEVPETIELEVTSTASGLLLWQGEVGE 145
Db 57 EIPQIGRSYLYTDNPDILKRVSG-----SR5-----NVFREFKTKADGILLNRG--DSP 105
QY 146 AQGGKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGEWHRVTLRREGRRGSIQVDGE 205
Db 106 MRPNDSDFISLGRDGLVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDQSGKITVDDY 165
QY 206 ELVSGRSPGNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLSARPGAPPPQ 265
Db 166 GARTKSPGMQLINGALYVGGMKELALHTNRQYMRGLVGCISHFTLST-----DY 218
QY 266 PDLQHRQAQAGANTRPC 282
Db 219 HISLVEDAVDGKNINTC 235

RESULT 5

US-10-073-865-121

; Sequence 121, Application US/10073865
; Publication No. US20030044904A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: RJZ09C1
; CURRENT APPLICATION NUMBER: US/10/073,865
; CURRENT FILING DATE: 2002-02-14
; Prior application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 154
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 121
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-073-865-121

Query Match 21.9%; Score 343; DB 14; Length 238;
Best Local Similarity 33.5%; Pred. No. 3.1e-22;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
QY 34 EHENPCOLREPLHGTC-----QTRCLCLPFGSGRCQSGHGHIAESDWHLESGGN 89
Db 11 ENAAHPC-VRAPCAHGSCSRKEGYDCDCPLGPEGLHCOKAITEAI-----56
QY 90 DAP-----GOYGAYFDDGFLAFPGHVFERSLPEVPETIELEVTSTASGLLLWQGEVGE 145
Db 57 EIPQIGRSYLYTDNPDILKRVSG-----SR5-----NVFREFKTKADGILLNRG--DSP 105
QY 146 AQGGKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGEWHRVTLRREGRRGSIQVDGE 205
Db 106 MRPNDSDFISLGRDGLVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDQSGKITVDDY 165
QY 206 ELVSGRSPGNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLSARPGAPPPQ 265
Db 166 GARTKSPGMQLINGALYVGGMKELALHTNRQYMRGLVGCISHFTLST-----DY 218
QY 266 PDLQHRQAQAGANTRPC 282
Db 219 HISLVEDAVDGKNINTC 235

RESULT 6

US-10-242-747-158

; Sequence 158, Application US/10242747
; Publication No. US20040005577A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ07C1
; CURRENT APPLICATION NUMBER: US/10/242,747
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,881
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 192
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-242-747-158

Query Match 21.9%; Score 343; DB 15; Length 238;
Best Local Similarity 33.5%; Pred. No. 3.1e-22;

```
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
QY 34 EHEENPCOLREPCLEGGTC-----QGTCLCLPFGSGPRCOQSGHGAESDWHLESGGN 89
Db 11 ENAAHPC-VRAPCAHGSCRPKEGYDCDPLGPEGLHCQKAIIEAL-----56
QY 90 DAP---GOYGAYFHDDGFLAFPGHVFPSRLPEVPTIELEVRTSTASGLLWQGVVEGE 145
Db 57 EIPQFIGRSYLTYPDILKRVSG---SRS-----NVFMFKTTAKDGLLWQGVVEGE 105
QY 146 AGQGDKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGEWHRVTLALREGRRGSIQVDGE 205
Db 106 MRPNDSFISLGRDGLVFSYNLGSGVASIMVNGSFNDGRWHRVKAVERDGGSGKITVDDY 165
QY 206 ELVSGRSPGNVAVNAKGVYIGGAPDVATLTGGRFSSGITGCVKNLVLSHARPGAPPPQ 265
Db 166 GARTGKSPGMRLQNLGALVYVGMKEIALHTNRQYMRGLVGCISHFTLST-----DY 218
QY 266 PLDQHRQAQAGANTRPC 282
Db 219 HISLVEDAVDGKNINTC 235

RESULT 7
US-09-978-249-10
; Sequence 10, Application US/09978249
; Patent No. US20020106780A1
; GENERAL INFORMATION:
; APPLICANT: Fisiella, et al.
; TITLE OF INVENTION: Extracellular Matrix Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT054P1
; CURRENT APPLICATION NUMBER: US/09/978,249
; PRIOR FILING DATE: 2001-10-17
; PRIOR APPLICATION NUMBER: PCT/US01/11643
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/198,123
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 375
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-978-249-10

Query Match 21.9%; Score 343; DB 9; Length 375;
Best Local Similarity 33.5%; Pred. No. 5.5e-22;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
QY 34 EHEENPCOLREPCLEGGTC-----QGTCLCLPFGSGPRCOQSGHGAESDWHLESGGN 89
Db 148 ENAAHPC-VRAPCAHGSCRPKEGYDCDPLGPEGLHCQKAIIEAL-----193
QY 90 DAP---GOYGAYFHDDGFLAFPGHVFPSRLPEVPTIELEVRTSTASGLLWQGVVEGE 145
Db 194 EIPQFIGRSYLTYPDILKRVSG---SRS-----NVFMFKTTAKDGLLWQGVVEGE 242
QY 146 AGQGDKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGEWHRVTLALREGRRGSIQVDGE 205
Db 243 MRPNDSFISLGRDGLVFSYNLGSGVASIMVNGSFNDGRWHRVKAVERDGGSGKITVDDY 302
QY 206 ELVSGRSPGNVAVNAKGVYIGGAPDVATLTGGRFSSGITGCVKNLVLSHARPGAPPPQ 265
Db 303 GARTGKSPGMRLQNLGALVYVGMKEIALHTNRQYMRGLVGCISHFTLST-----DY 355
QY 266 PLDQHRQAQAGANTRPC 282
Db 356 HISLVEDAVDGKNINTC 372

RESULT 8
US-09-764-853-541
; Sequence 541, Application US/09764853
```

```
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P0206
; CURRENT APPLICATION NUMBER: US/09/764,853
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 541
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-853-541

Query Match 21.9%; Score 343; DB 9; Length 432;
Best Local Similarity 33.5%; Pred. No. 6.6e-22;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
QY 34 EHEENPCOLREPCLEGGTC-----QGTCLCLPFGSGPRCOQSGHGAESDWHLESGGN 89
Db 205 ENAAHPC-VRAPCAHGSCRPKEGYDCDPLGPEGLHCQKAIIEAL-----250
QY 90 DAP---GOYGAYFHDDGFLAFPGHVFPSRLPEVPTIELEVRTSTASGLLWQGVVEGE 145
Db 251 EIPQFIGRSYLTYPDILKRVSG---SRS-----NVFMFKTTAKDGLLWQGVVEGE 299
QY 146 AGQGDKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGEWHRVTLALREGRRGSIQVDGE 205
Db 300 MRPNDSFISLGRDGLVFSYNLGSGVASIMVNGSFNDGRWHRVKAVERDGGSGKITVDDY 359
QY 206 ELVSGRSPGNVAVNAKGVYIGGAPDVATLTGGRFSSGITGCVKNLVLSHARPGAPPPQ 265
Db 360 GARTGKSPGMRLQNLGALVYVGMKEIALHTNRQYMRGLVGCISHFTLST-----DY 412
QY 266 PLDQHRQAQAGANTRPC 282
Db 413 HISLVEDAVDGKNINTC 429

RESULT 9
US-09-764-898-179
; Sequence 179, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P0201
; CURRENT APPLICATION NUMBER: US/09/764,898
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 311
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 179
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-898-179

Query Match 21.9%; Score 343; DB 9; Length 432;
Best Local Similarity 33.5%; Pred. No. 6.6e-22;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
QY 34 EHEENPCOLREPCLEGGTC-----QGTCLCLPFGSGPRCOQSGHGAESDWHLESGGN 89
Db 205 ENAAHPC-VRAPCAHGSCRPKEGYDCDPLGPEGLHCQKAIIEAL-----250
QY 90 DAP---GOYGAYFHDDGFLAFPGHVFPSRLPEVPTIELEVRTSTASGLLWQGVVEGE 145
Db 251 EIPQFIGRSYLTYPDILKRVSG---SRS-----NVFMFKTTAKDGLLWQGVVEGE 299
QY 146 AGQGDKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGEWHRVTLALREGRRGSIQVDGE 205
```

Db 300 MRPNDFISLGRDGLVFNLSGVASIMVNGSFNDGRHVRKAVRDGSGKITVDDY 359
QY 206 ELVSGRSPGNVAVNAKSVYIGGAPDVATLTGGRFSSGITGCVKNLVLSHARPGAPPPQ 265
Db 360 GARTGKSPGMQNLINGALVYVGMKEIALHTNRQYMRGLVGCISHFTLST-----DY 412
QY 266 PLDLOHRAQAAGANTRPC 282
Db 413 HISLVEDAVDGKNINTC 429

RESULT 10
US-09-764-881-102
; Sequence 102, Application US/09764881
; Publication No. US20030125246A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT207
; CURRENT APPLICATION NUMBER: US/09/764,881
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 192
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 102
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-881-102

Query Match 21.9%; Score 343; DB 10; Length 432;
Best Local Similarity 33.5%; Pred. No. 6.6e-22;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

QY 34 EHEENPCOLREPCPLHGTC-----QGTCLCLPFGSPRCQOQSGHGIABSDWHLESGGN 89
Db 205 ENAAHPC-VRAPCAHGSCRPKEGYDCDCPLGFEGLHCOKAIEAI----- 250
QY 90 DAP-----QOYGAYFHDDGFLAPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVVEGE 145
Db 300 MRPNDFISLGRDGLVFNLSGVASIMVNGSFNDGRHVRKAVRDGSGKITVDDY 359
QY 206 ELVSGRSPGNVAVNAKSVYIGGAPDVATLTGGRFSSGITGCVKNLVLSHARPGAPPPQ 265
Db 360 GARTGKSPGMQNLINGALVYVGMKEIALHTNRQYMRGLVGCISHFTLST-----DY 412
QY 266 PLDLOHRAQAAGANTRPC 282
Db 413 HISLVEDAVDGKNINTC 429

RESULT 11
US-10-073-865-78
; Sequence 78, Application US/10073865
; Publication No. US20030044904A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P209C1
; CURRENT APPLICATION NUMBER: US/10/073,865
; CURRENT FILING DATE: 2002-02-14
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 154
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens

US-10-073-865-78

Query Match 21.9%; Score 343; DB 14; Length 432;
Best Local Similarity 33.5%; Pred. No. 6.6e-22;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

QY 34 EHEENPCOLREPCPLHGTC-----QGTCLCLPFGSPRCQOQSGHGIABSDWHLESGGN 89
Db 205 ENAAHPC-VRAPCAHGSCRPKEGYDCDCPLGFEGLHCOKAIEAI----- 250
QY 90 DAP-----QOYGAYFHDDGFLAPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVVEGE 145
Db 251 EIPQFIGRSYLTYPNDILKRVSG-----NVFMRFKTTAKDGLLWRG--DSP 299
QY 146 AQGKDFISLGRDGLVFNLSGVASIMVNGSFNDGRHVRKAVRDGSGKITVDDY 359
Db 300 MRPNDFISLGRDGLVFNLSGVASIMVNGSFNDGRHVRKAVRDGSGKITVDDY 359
QY 206 ELVSGRSPGNVAVNAKSVYIGGAPDVATLTGGRFSSGITGCVKNLVLSHARPGAPPPQ 265
Db 360 GARTGKSPGMQNLINGALVYVGMKEIALHTNRQYMRGLVGCISHFTLST-----DY 412
QY 266 PLDLOHRAQAAGANTRPC 282
Db 413 HISLVEDAVDGKNINTC 429

RESULT 12
US-10-242-747-102
; Sequence 102, Application US/10242747
; Publication No. US20040005577A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT207C1
; CURRENT APPLICATION NUMBER: US/10/242,747
; CURRENT FILING DATE: 2002-09-13
; Prior APPLICATION NUMBER: 09/764,881
; Prior FILING DATE: 2001-01-17
; Prior APPLICATION NUMBER: 60/179,065
; Prior FILING DATE: 2000-01-31
; Prior APPLICATION NUMBER: 60/180,628
; Prior FILING DATE: 2000-02-04
; Prior APPLICATION NUMBER: 60/214,886
; Prior FILING DATE: 2000-06-28
; Prior APPLICATION NUMBER: 60/217,487
; Prior FILING DATE: 2000-07-11
; Prior APPLICATION NUMBER: 60/225,758
; Prior FILING DATE: 2000-08-14
; Prior APPLICATION NUMBER: 60/220,963
; Prior FILING DATE: 2000-07-26
; Prior APPLICATION NUMBER: 60/217,496
; Prior FILING DATE: 2000-07-11
; Prior APPLICATION NUMBER: 60/225,447
; Prior FILING DATE: 2000-08-14
; Prior APPLICATION NUMBER: 60/218,290
; Prior FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 192
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 102
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-242-747-102

Query Match 21.9%; Score 343; DB 15; Length 432;
Best Local Similarity 33.5%; Pred. No. 6.6e-22;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

QY 34 EHEENPCOLREPCPLHGTC-----QGTCLCLPFGSPRCQOQSGHGIABSDWHLESGGN 89
Db 205 ENAAHPC-VRAPCAHGSCRPKEGYDCDCPLGFEGLHCOKAIEAI----- 250

Thu Mar 11 09:25:23 2004

Db 101 CGRRPRVLGESPVHTVNLKEPLVVGAPDFSKLARAANVSSGFDGAIQLVSLGGRQLL 160
QY 31 -----DLCEHEENPCQLR--EPCILHGGTCQGTTR-----CLCLPGFSGPRCQOQSGHG 75
Db 161 TPEHVLROVDVTSFAGHPCTRASGHPCPLNGASCVPREAAVCLCPGGFSGPHCEKGL--- 217
QY 76 IABSDWHLEGSGCNDAPGOYGAYFHDDGFLAPPGHVFSSRLPEVPET-----IELEV 127
Db 218 -----VEKSAG-----DVTLPADGRTFVEYLNATSEKALQSNHFELS 258
QY 128 RTSTASGLLLMQGVVEVBAGQKDFISLQDGHVLFYRIQLGSGEARLVSEDPINDGEWH 187
Db 259 RTEATQGLVMS-----GKATERADYVALAIVDGHLQLSYNLGSPVVLRSVPVNTNRWL 314
QY 188 RVTALREGRRSIQVDGEELVSGRSGPNVAVNAGSVYIGCAPD--VATLTGGRFSSGI 245
Db 315 RVVAHREQREGSLQVNEAPVTGSSPLGATQDLDGALWLGGLPELPVGPALPKAYGTGF 374
QY 246 TGCVKNLVLHSARPGAPPPQPLDLQHRQAQAGANTRPCPS 284
Db 375 VGCLRDVVVGR-----HPLHLLLEDVTKPELRPCPT 405

Search completed: March 9, 2004, 17:25:15
Job time : 21.844 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:16:23 ; Search time 11.1326 Seconds
(without alignments)
1317.011 Million cell updates/sec

Title: US-10-006-011a-9
Perfect score: 1566
Sequence: 1 CERPCQHGATCMGAYEF.....QPLDLQRAQAGNTRPCPS 284

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	333.5	21.3	1940	2	US-08-644-271-30
2	333.5	21.3	1940	4	US-09-077-955-34
3	330	21.1	294	4	US-09-077-955-29
4	330	21.1	338	4	US-09-077-955-28
5	330	21.1	390	4	US-09-077-955-27
6	330	21.1	440	4	US-09-077-955-26
7	330	21.1	456	4	US-09-077-955-25
8	330	21.1	492	2	US-08-644-271-32
9	330	21.1	492	4	US-09-077-955-36
10	326	20.8	256	4	US-09-077-955-30
11	261	16.7	216	4	US-09-077-955-31
12	253.5	16.2	1525	3	US-09-191-647-2
13	253.5	16.2	1525	3	US-09-540-245A-2
14	253.5	16.2	1525	3	US-09-540-153-2
15	249	15.9	735	3	US-09-191-647-9
16	249	15.9	735	3	US-09-540-245A-9
17	249	15.9	735	3	US-09-540-153-9
18	245.5	15.7	1529	4	US-09-312-283C-396
19	239	15.3	1523	4	US-09-182-024A-2
20	226.5	14.5	716	4	US-09-312-283C-183
21	226.5	14.5	771	3	US-09-188-930-183
22	223.5	14.3	1130	2	US-08-460-309-2
23	223.5	14.3	1130	2	US-08-125-077-2
24	223.5	14.3	1130	6	544158-2
25	223.5	14.3	3088	4	US-09-562-702A-8
26	223.5	14.3	3089	4	US-09-562-702A-4
27	223.5	14.3	3110	4	US-09-562-702A-2

Query Match 21.3%; Score 333.5; DB 2; Length 1940;

ALIGNMENTS

RESULT 1
US-08-644-271-30
; Sequence 30, Application US/08644271
; Patent No. 5814478
; GENERAL INFORMATION:
; APPLICANT: Valenzuela, et al.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
; TITLE OF INVENTION: AND LIGANDS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/644,271
; FILING DATE: 10-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 60/008,657
; FILING DATE: 15-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Cobert, Robert J
; REGISTRATION NUMBER: 36,108
; REFERENCE/DOCKET NUMBER: REG 195A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 914-345-7400
; TELEFAX: 914-345-7721
; TELEX:
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1940 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Rat Agrin
; LOCATION: 1...1940
; OTHER INFORMATION:
US-08-644-271-30

Sequence 6, Appli
Sequence 7, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 12, Appli
Sequence 10, Appli
Sequence 2, Appli
Sequence 4, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 2, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 2, Appli
Sequence 7, Appli
Sequence 7, Appli
Sequence 2, Appli

Best Local Similarity 36.0%; Pred. No. 9.2e-22;
Matches 94; Conservative 28; Mismatches 108; Indels 31; Gaps 10;

QY 1 CERQPCOHGATCMPAGEYFQCLCRDGFKGLDCEHENPCQLREPCCLHGGTCQ-----GT 55
Db 1444 CLPNPCHGALCOALEAGMFLCQCPRGFGPTCADEKSPCQ-PNPGHGAAPCRVLSGGA 1502

QY 56 RCLCLFGSGPRCQOQSGHGAESDWHLEGGSGNDAPQGYGAYFHDGFLAPPG-HVFSR 114
Db 1503 KCECLGRSGTFCQT-----VLETAGSR-----PFLADFNFGFSYLEKGLHTFER 1547

QY 115 SLPEVPETTELEVRTSTAGLLWQGVGEAGQKDFISLQDGHVFRYQLGSGEAR 174
Db 1548 DLGE-KMALEMVFLARGPSGLLYNGQTD--GKG-DFVSLAHNRHLEFCYDLGKGA 1603

QY 175 LVSEDPINDGEWHRTVLRGREGSGTQVQDGBELVSGRSPG-----PNVAVNAKGSVYIGGA 230
Db 1604 IRSKEPIALGTWVRVFLERNRKGALQVGDGPRVLGSEPKRKVPHTMLNKEPLYIGGA 1663

QY 231 PDVATLT-GRFSSGIGTCVK 250
Db 1664 PDFSKLARGAAGVSSGFSGVIIQ 1684

RESULT 2
US-09-077-955-34
; Sequence 34, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; EARLIER FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 34
; LENGTH: 1940
; TYPE: PRT
; ORGANISM: Rattus sp.
US-09-077-955-34

Query Match 21.3%; Score 333.5; DB 4; Length 1940;
Best Local Similarity 36.0%; Pred. No. 9.2e-22;
Matches 94; Conservative 28; Mismatches 108; Indels 31; Gaps 10;

QY 1 CERQPCOHGATCMPAGEYFQCLCRDGFKGLDCEHENPCQLREPCCLHGGTCQ-----GT 55
Db 1444 CLPNPCHGALCOALEAGMFLCQCPRGFGPTCADEKSPCQ-PNPGHGAAPCRVLSGGA 1502

QY 56 RCLCLFGSGPRCQOQSGHGAESDWHLEGGSGNDAPQGYGAYFHDGFLAPPG-HVFSR 114
Db 1503 KCECLGRSGTFCQT-----VLETAGSR-----PFLADFNFGFSYLEKGLHTFER 1547

QY 115 SLPEVPETTELEVRTSTAGLLWQGVGEAGQKDFISLQDGHVFRYQLGSGEAR 174
Db 1548 DLGE-KMALEMVFLARGPSGLLYNGQTD--GKG-DFVSLAHNRHLEFCYDLGKGA 1603

QY 175 LVSEDPINDGEWHRTVLRGREGSGTQVQDGBELVSGRSPG-----PNVAVNAKGSVYIGGA 230
Db 1604 IRSKEPIALGTWVRVFLERNRKGALQVGDGPRVLGSEPKRKVPHTMLNKEPLYIGGA 1663

QY 231 PDVATLT-GRFSSGIGTCVK 250
Db 1664 PDFSKLARGAAGVSSGFSGVIIQ 1684

Query Match 21.3%; Score 333.5; DB 4; Length 1940;
Best Local Similarity 36.0%; Pred. No. 9.2e-22;
Matches 94; Conservative 28; Mismatches 108; Indels 31; Gaps 10;

QY 1 CERQPCOHGATCMPAGEYFQCLCRDGFKGLDCEHENPCQLREPCCLHGGTCQ-----GT 55
Db 1444 CLPNPCHGALCOALEAGMFLCQCPRGFGPTCADEKSPCQ-PNPGHGAAPCRVLSGGA 1502

QY 56 RCLCLFGSGPRCQOQSGHGAESDWHLEGGSGNDAPQGYGAYFHDGFLAPPG-HVFSR 114
Db 1503 KCECLGRSGTFCQT-----VLETAGSR-----PFLADFNFGFSYLEKGLHTFER 1547

QY 115 SLPEVPETTELEVRTSTAGLLWQGVGEAGQKDFISLQDGHVFRYQLGSGEAR 174
Db 1548 DLGE-KMALEMVFLARGPSGLLYNGQTD--GKG-DFVSLAHNRHLEFCYDLGKGA 1603

QY 175 LVSEDPINDGEWHRTVLRGREGSGTQVQDGBELVSGRSPG-----PNVAVNAKGSVYIGGA 230
Db 1604 IRSKEPIALGTWVRVFLERNRKGALQVGDGPRVLGSEPKRKVPHTMLNKEPLYIGGA 1663

QY 231 PDVATLT-GRFSSGIGTCVK 250
Db 1664 PDFSKLARGAAGVSSGFSGVIIQ 1684

RESULT 3
US-09-077-955-29
; Sequence 29, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; EARLIER FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 29
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-077-955-29

Query Match 21.1%; Score 330; DB 4; Length 294;
Best Local Similarity 31.3%; Pred. No. 1.5e-22;
Matches 87; Conservative 34; Mismatches 99; Indels 58; Gaps 8;

QY 31 DLCEHENPCQLR--EPCLHGGTCQGT-----CLCLPGFSGPRCQOQSGHGAESDWHLE 84
Db 50 DVTSPAGHPCCTASGHPCFLNGASCVPREAAVCLCPGFGSGPHCEKGL-----VE 99

QY 85 GSGGNDAPQGYGAYFHDGFLAPPGHVFSRSLPEVET-----ILEVR 128
Db 100 KSAG-----DVTFLAFDGRTFVEYLAANVTELANEIPVEKALQSNHFLSLR 147

QY 129 TSTASGILLWQGVGEAGQKDFISLQDGHVFRYQLGSGEARLIVSDPINDGEWHR 188
Db 148 TEATQGLVMS---GRATERADYVALAIVDGHQLSYNLGSPVVLRSVPTVNTKWL 203

QY 189 VTALREGRRGSIQVDGELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGIT 246
Db 204 VVAHREQREGSLQVNEAPVTGSSPLGATQLDITDGLMLGGLPELPVGPALPKAYGTGFV 263

QY 247 GCVKNLVLHSAAPGAPPPQPLDQHRAGAGANTRPSPS 284
Db 264 GCLRDVVYGR-----HPLHLEDAVTKPELRPCPT 293

RESULT 4
US-09-077-955-28
; Sequence 28, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; EARLIER FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 28
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-077-955-28

Query Match 21.1%; Score 330; DB 4; Length 294;
Best Local Similarity 31.3%; Pred. No. 1.5e-22;
Matches 87; Conservative 34; Mismatches 99; Indels 58; Gaps 8;

QY 31 DLCEHENPCQLR--EPCLHGGTCQGT-----CLCLPGFSGPRCQOQSGHGAESDWHLE 84
Db 50 DVTSPAGHPCCTASGHPCFLNGASCVPREAAVCLCPGFGSGPHCEKGL-----VE 99

QY 85 GSGGNDAPQGYGAYFHDGFLAPPGHVFSRSLPEVET-----ILEVR 128
Db 100 KSAG-----DVTFLAFDGRTFVEYLAANVTELANEIPVEKALQSNHFLSLR 147

QY 129 TSTASGILLWQGVGEAGQKDFISLQDGHVFRYQLGSGEARLIVSDPINDGEWHR 188
Db 148 TEATQGLVMS---GRATERADYVALAIVDGHQLSYNLGSPVVLRSVPTVNTKWL 203

QY 189 VTALREGRRGSIQVDGELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGIT 246
Db 204 VVAHREQREGSLQVNEAPVTGSSPLGATQLDITDGLMLGGLPELPVGPALPKAYGTGFV 263

QY 247 GCVKNLVLHSAAPGAPPPQPLDQHRAGAGANTRPSPS 284
Db 264 GCLRDVVYGR-----HPLHLEDAVTKPELRPCPT 293

Query Match 21.1%; Score 330; DB 4; Length 338;
 Best Local Similarity 31.3%; Pred. No. 1.9e-22;
 Matches 87; Conservative 34; Mismatches 99; Indels 58; Gaps 8;

QY 31 DLCEHENPCQLR--EPC LHGGTCQGR-----CLCLPGFSGPRCQCGSGHGIAESDWHLE 84
 Db 94 DVTSFAGHPCTRASGHPCLNGASCVPREAAVCLCPGFSGPHCEKGL-----VE 143
 QY 85 GSGNDAPGQYGAFHDDGFLAFPGHVSRLPVPET-----ILEVR 128
 Db 144 KSAG-----DVTTLAFDGRTFVEYLNVAVTESELANEIPVEKALQSNHFFELSRL 191
 QY 129 TSTASGLLLWQGVGEAGQKDFISLQDGHVFRYQLGSGEARLVSDPINDGEWHR 188
 Db 192 TEATQGLVWS-----GKATERADYVALAIVDGHQLSYNLGSPVVLSTVPVNTNRMLR 247
 QY 189 VTALREGRGSIQVDGEBELVSGRSPGNVAVNAKSVYIGAPD--VATLTGGRFSSGIT 246
 Db 248 VVAHREQREGSLQVNEAPVTGSSPLGATQDLDGALWGLGPELPGVPALPKAYGTGFV 307
 QY 247 GCVKNLVLSARPAPPPQPLDQRAQAGANTRPCPS 284
 Db 308 GCLRDVVVGR-----HPLHLEDAVTKPELRPCPT 337

RESULT 5

US-09-077-955-27
 ; Sequence 27, Application US/09077955A
 ; Patent No. 6413740
 ; GENERAL INFORMATION:
 ; APPLICANT: Valenzuela et al., David M.
 ; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
 ; FILE REFERENCE: REG195-B-PCT-US
 ; CURRENT APPLICATION NUMBER: US/09/077,955A
 ; EARLIER FILING DATE: 1998-09-10
 ; EARLIER APPLICATION NUMBER: PCT/US96/20696
 ; EARLIER FILING DATE: 1996-12-13
 ; EARLIER APPLICATION NUMBER: 08/644,271
 ; EARLIER FILING DATE: 1996-05-10
 ; EARLIER APPLICATION NUMBER: 60/008,657
 ; EARLIER FILING DATE: 1995-12-15
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 27
 ; LENGTH: 390
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-077-955-27

Query Match 21.1%; Score 330; DB 4; Length 390;
 Best Local Similarity 31.3%; Pred. No. 2.3e-22;
 Matches 87; Conservative 34; Mismatches 99; Indels 58; Gaps 8;

QY 31 DLCEHENPCQLR--EPC LHGGTCQGR-----CLCLPGFSGPRCQCGSGHGIAESDWHLE 84
 Db 146 DVTSFAGHPCTRASGHPCLNGASCVPREAAVCLCPGFSGPHCEKGL-----VE 195
 QY 85 GSGNDAPGQYGAFHDDGFLAFPGHVSRLPVPET-----ILEVR 128
 Db 196 KSAG-----DVTTLAFDGRTFVEYLNVAVTESELANEIPVEKALQSNHFFELSRL 243
 QY 129 TSTASGLLLWQGVGEAGQKDFISLQDGHVFRYQLGSGEARLVSDPINDGEWHR 188
 Db 244 TEATQGLVWS-----GKATERADYVALAIVDGHQLSYNLGSPVVLSTVPVNTNRMLR 299
 QY 189 VTALREGRGSIQVDGEBELVSGRSPGNVAVNAKSVYIGAPD--VATLTGGRFSSGIT 246
 Db 300 VVAHREQREGSLQVNEAPVTGSSPLGATQDLDGALWGLGPELPGVPALPKAYGTGFV 359
 QY 247 GCVKNLVLSARPAPPPQPLDQRAQAGANTRPCPS 284
 Db 360 GCLRDVVVGR-----HPLHLEDAVTKPELRPCPT 389

RESULT 6

US-09-077-955-26
 ; Sequence 26, Application US/09077955A
 ; Patent No. 6413740
 ; GENERAL INFORMATION:
 ; APPLICANT: Valenzuela et al., David M.
 ; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
 ; FILE REFERENCE: REG195-B-PCT-US
 ; CURRENT APPLICATION NUMBER: US/09/077,955A
 ; EARLIER FILING DATE: 1998-09-10
 ; EARLIER APPLICATION NUMBER: PCT/US96/20696
 ; EARLIER FILING DATE: 1996-12-13
 ; EARLIER APPLICATION NUMBER: 08/644,271
 ; EARLIER FILING DATE: 1996-05-10
 ; EARLIER APPLICATION NUMBER: 60/008,657
 ; EARLIER FILING DATE: 1995-12-15
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 26
 ; LENGTH: 440
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-077-955-26

Query Match 21.1%; Score 330; DB 4; Length 440;
 Best Local Similarity 31.3%; Pred. No. 2.7e-22;
 Matches 87; Conservative 34; Mismatches 99; Indels 58; Gaps 8;

QY 31 DLCEHENPCQLR--EPC LHGGTCQGR-----CLCLPGFSGPRCQCGSGHGIAESDWHLE 84
 Db 196 DVTSFAGHPCTRASGHPCLNGASCVPREAAVCLCPGFSGPHCEKGL-----VE 245
 QY 85 GSGNDAPGQYGAFHDDGFLAFPGHVSRLPVPET-----ILEVR 128
 Db 246 KSAG-----DVTTLAFDGRTFVEYLNVAVTESELANEIPVEKALQSNHFFELSRL 293
 QY 129 TSTASGLLLWQGVGEAGQKDFISLQDGHVFRYQLGSGEARLVSDPINDGEWHR 188
 Db 294 TEATQGLVWS-----GKATERADYVALAIVDGHQLSYNLGSPVVLSTVPVNTNRMLR 349
 QY 189 VTALREGRGSIQVDGEBELVSGRSPGNVAVNAKSVYIGAPD--VATLTGGRFSSGIT 246
 Db 350 VVAHREQREGSLQVNEAPVTGSSPLGATQDLDGALWGLGPELPGVPALPKAYGTGFV 409
 QY 247 GCVKNLVLSARPAPPPQPLDQRAQAGANTRPCPS 284
 Db 410 GCLRDVVVGR-----HPLHLEDAVTKPELRPCPT 439

RESULT 7

US-09-077-955-25
 ; Sequence 25, Application US/09077955A
 ; Patent No. 6413740
 ; GENERAL INFORMATION:
 ; APPLICANT: Valenzuela et al., David M.
 ; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
 ; FILE REFERENCE: REG195-B-PCT-US
 ; CURRENT APPLICATION NUMBER: US/09/077,955A
 ; EARLIER FILING DATE: 1998-09-10
 ; EARLIER APPLICATION NUMBER: PCT/US96/20696
 ; EARLIER FILING DATE: 1996-12-13
 ; EARLIER APPLICATION NUMBER: 08/644,271
 ; EARLIER FILING DATE: 1996-05-10
 ; EARLIER APPLICATION NUMBER: 60/008,657
 ; EARLIER FILING DATE: 1995-12-15
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 25
 ; LENGTH: 456
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-077-955-25

```

Query Match      21.1%; Score 330; DB 4; Length 456;
Best Local Similarity 31.3%; Pred. No. 2.8e-22;
Matches 87; Conservative 34; Mismatches 99; Indels 58; Gaps 8;

QY 31 DLCEHENPCQLR--EPLCHGTCQGR-----CLCLPGSGPRCCQCGSGHGAESDWHLE 84
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 212 DVTSFAGHPCPTRASGHPCPLNGASCVPREAAVCLCPGSGSPHCEKGL-----VE 261
QY 85 GSGNDAPGQYGAYFDDGFLAPPGHVSRSLEVPET-----TELEVR 128
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 262 KSAG-----DVTDLAFDGRTFVEYLNVAVTESELANEIPVEKALQSNHFLSLR 309
QY 129 TSTASGLLLMOGVEVGEAGQKDFISLGLQDGHVRYOLGSGEARLVSEDPINDGEWHR 188
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 310 TEATQGLVLS-----GKATERADYVALAIVDGHQLSYNLGSPQVVLRSVTVNTNRWLR 365
QY 189 VTALREGRRGSIQVDGEELVSGRSPGNVAVNAKSVYIGGAPD--VATLTGGRFSSGIT 246
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 366 VVAHQREGSLQVNEAPVTGSSPLGATQDLDGALMLGGLPELPVGPALPKAYGTGFV 425
QY 247 GCVKNLVLSARPGAPPPQPLDLOHRAQAQANTRPCPS 284
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 426 GCLRDVVVGR-----HPLHLEDAVTKPBLRPOPT 455

```

```

RESULT 8
US-08-644-271-32
; Sequence 32, Application US/08644271
; Patent No. 5814478
; GENERAL INFORMATION:
; APPLICANT: Valenzuela, et al.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/644,271
; FILING DATE: 10-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 60/008,657
; FILING DATE: 15-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Cobert, Robert J
; REGISTRATION NUMBER: 36,108
; REFERENCE/DOCKET NUMBER: REG 195A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 914-345-7400
; TELEFAX: 914-345-7721
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 492 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; NAME/KEY: Human Agrin
; LOCATION: 1...492

```

```

OTHER INFORMATION:
US-08-644-271-32
Query Match      21.1%; Score 330; DB 2; Length 492;
Best Local Similarity 31.3%; Pred. No. 3.1e-22;
Matches 87; Conservative 34; Mismatches 99; Indels 58; Gaps 8;

QY 31 DLCEHENPCQLR--EPLCHGTCQGR-----CLCLPGSGPRCCQCGSGHGAESDWHLE 84
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 248 DVTSFAGHPCPTRASGHPCPLNGASCVPREAAVCLCPGSGSPHCEKGL-----VE 297
QY 85 GSGNDAPGQYGAYFDDGFLAPPGHVSRSLEVPET-----TELEVR 128
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 298 KSAG-----DVTDLAFDGRTFVEYLNVAVTESELANEIPVEKALQSNHFLSLR 345
QY 129 TSTASGLLLMOGVEVGEAGQKDFISLGLQDGHVRYOLGSGEARLVSEDPINDGEWHR 188
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 346 TEATQGLVLS-----GKATERADYVALAIVDGHQLSYNLGSPQVVLRSVTVNTNRWLR 401
QY 189 VTALREGRRGSIQVDGEELVSGRSPGNVAVNAKSVYIGGAPD--VATLTGGRFSSGIT 246
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 402 VVAHQREGSLQVNEAPVTGSSPLGATQDLDGALMLGGLPELPVGPALPKAYGTGFV 461
QY 247 GCVKNLVLSARPGAPPPQPLDLOHRAQAQANTRPCPS 284
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 462 GCLRDVVVGR-----HPLHLEDAVTKPBLRPOPT 491

```

```

RESULT 9
US-09-077-955-36
; Sequence 36, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 36
; TYPE: PRT
; LENGTH: 492
; ORGANISM: Homo sapiens
US-09-077-955-36
Query Match      21.1%; Score 330; DB 4; Length 492;
Best Local Similarity 31.3%; Pred. No. 3.1e-22;
Matches 87; Conservative 34; Mismatches 99; Indels 58; Gaps 8;

QY 31 DLCEHENPCQLR--EPLCHGTCQGR-----CLCLPGSGPRCCQCGSGHGAESDWHLE 84
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 248 DVTSFAGHPCPTRASGHPCPLNGASCVPREAAVCLCPGSGSPHCEKGL-----VE 297
QY 85 GSGNDAPGQYGAYFDDGFLAPPGHVSRSLEVPET-----TELEVR 128
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 298 KSAG-----DVTDLAFDGRTFVEYLNVAVTESELANEIPVEKALQSNHFLSLR 345
QY 129 TSTASGLLLMOGVEVGEAGQKDFISLGLQDGHVRYOLGSGEARLVSEDPINDGEWHR 188
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 346 TEATQGLVLS-----GKATERADYVALAIVDGHQLSYNLGSPQVVLRSVTVNTNRWLR 401
QY 189 VTALREGRRGSIQVDGEELVSGRSPGNVAVNAKSVYIGGAPD--VATLTGGRFSSGIT 246
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 402 VVAHQREGSLQVNEAPVTGSSPLGATQDLDGALMLGGLPELPVGPALPKAYGTGFV 461
QY 247 GCVKNLVLSARPGAPPPQPLDLOHRAQAQANTRPCPS 284

```

```
Db 462 GCLRDVVVGR-----HPLHLEDAVTKPELRPCPT 491
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-077-955-31

Query Match 16.7%; Score 261; DB 4; Length 216;
Best Local Similarity 32.8%; Pred. No. 2.4e-16;
Matches 65; Conservative 27; Mismatches 76; Indels 30; Gaps 4;

QY 105 LAPPGHVFSRSLPEVPET-----IELEVRTSTASGLLLMQGVGEAGQ 148
Db 30 LAFDGRFTVEYLNATESELANEIPVEKALQSNHFLSLRTEATCGVLNLS-----GKATE 85
QY 149 GKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGEWHRVTLRREGRRSGIQVDGBELY 208
Db 86 RADYVALAIVDGHQLQSLNLSGQPWVLRSTVFNTRWLRVVAHREOREGSLQVNEAPV 145
QY 209 SGRSPGPNVAVNAKGSVYIGAPD--VATLTGGRFSSGITCGKVLVLSARPAPPPQP 266
Db 146 TGSSPLGATQDLDGALWLGGLPELPVGPALPKAYGTGFGCLRDVVVGR-----HP 197
QY 267 LDQHRAGAGANTREPCS 284
Db 198 LHLLEDAVTKPELRPCPT 215

RESULT 12
US-09-191-647-2
; Sequence 2, Application US/09191647
; Patent No. 6046015
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/191,647
; CURRENT FILING DATE: 1998-11-13
; EARLIER APPLICATION NUMBER: 60/065,544
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: 60/081,057
; EARLIER FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: human
US-09-191-647-2

Query Match 16.2%; Score 253.5; DB 3; Length 1525;
Best Local Similarity 24.8%; Pred. No. 1.6e-14;
Matches 92; Conservative 48; Mismatches 100; Indels 131; Gaps 18;

QY 1 CERQPCQHGATC-MPAGEYE--FQCLCRDGFVKDICE-----HEENPC----- 40
Db 957 CISNPKCKGGTCHLKEGSEDFWCICADGFECEVNVDDCEDNCCENNSTCVDGINNY 1016
QY 41 -----QLREPCGLHGTC-----QSTRCLCLPGFSGP----- 66
Db 1017 TCLCPPEVTGELCEEKLDPCADLNPCQHDSKCLITPKFKCDCTPGVYGEHCDIDFDDC 1076
QY 67 ---RQOQSGHGIAESDWHL---EGSGG-----NDA----- 91
Db 1077 QDNCKNGAHTCTDAVNGYTCICPEGYSGLFCFESPPMVLPRTPSDFNDFCQNGAQCVRI 1136
QY 92 -----PGOYGA-----YFDDGFLAFPGHVFPSLSPEVPET--IELEVRTSTAS 133
Db 1137 NEPICQCLPGYQGEKCEKLVSNFINKESYLQIP-----SAKVRPQTNTLTQATDEDS 1190
QY 134 GLLLMQGVGEAGQKDFISLGLQDGHVFRYQLGSGEARLV--SEDPINDGEWHRVTL 192
Db 1191 GILLYKG-----DKDHIAYELRYGRVRSYDTGSHPASAIYSVETINDGNFIVELL 1242

RESULT 11
US-09-077-955-31
; Sequence 31, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 31
; LENGTH: 216

Query Match 20.8%; Score 326; DB 4; Length 256;
Best Local Similarity 32.1%; Pred. No. 3e-22; Indels 56; Gaps 7;
Matches 84; Conservative 32; Mismatches 90; Indels 56; Gaps 7;

QY 45 PCLHGTCQGTR-----CLCLPGFSGRCQQSGHGHIAESDWHLEGGGNDAPGQYGYF 100
Db 28 PCLNGASCVPREAAVYCLCPGFGSPHCEKGL-----VEKSAG----- 65
QY 101 DQGLAFPGHVSRSLSPEVPET-----IELEVRTSTASGLLLMQGVYEVG 144
Db 66 DVDTLAFDGRFTVEYLNATESELANEIPVEKALQSNHFLSLRTEATCGVLNLS-----G 121
QY 145 EAQGGKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGEWHRVTLRREGRRSGIQVDG 204
Db 122 KATERADYVALAIVDGHQLQSLNLSGQPWVLRSTVFNTRWLRVVAHREOREGSLQVGN 181
QY 205 EELVSGSPGPNVAVNAKGSVYIGAPD--VATLTGGRFSSGITCGKVLVLSARPAP 262
Db 182 EAPVTGSSPLGATQDLDGALWLGGLPELPVGPALPKAYGTGFGCLRDVVVGR----- 235
QY 263 PPQPLDQHRAGAGANTREPCS 284
Db 236 --HPLHLEDAVTKPELRPCPT 255

RESULT 10
US-09-077-955-30
; Sequence 30, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 30
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-077-955-30
```



```
QY 193 REGRGSIQVDG--BELVSGRSPGNVAVNAKGSVYIGAP---DVATL--TGGRFSSGI 245
Db 1243 ALDQSLSLVDGNGPKIITNLS--KQSTLNFDSPLYVGMFGKSNVSLRQAPQNGTSF 1300
QY 246 TGCVRNVLVHS 256
Db 1301 HGCIRNLYINS 1311

RESULT 13
US-09-540-245A-2
; Sequence 2, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; CURRENT APPLICATION NUMBER: US/09/540,245A
; CURRENT FILING DATE: 2000-03-31
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: human
US-09-540-245A-2

Query Match 16.2%; Score 253.5; DB 3; Length 1525;
Best Local Similarity 24.8%; Pred. No. 1.6e-14;
Matches 92; Conservative 48; Mismatches 100; Indels 131; Gaps 18;

QY 1 CERPCQCHGATC-MPAGEYE-FQCLCRDGFPGDLCE-----HEENPC----- 40
Db 957 CISNPKCHGGTCHLKEGEDGFWCICADGFECEVNVDDCEDNCCNNSTCVDGINNY 1016
QY 41 -----QLREPCCLHGTC-----QGTCLCLPGFSGP----- 66
Db 1017 TCLCPPEYTGELCEKLDLFCQDLNFCQHDSKILTPKFKDCCTPGVYGEHCHDIDFDC 1076
QY 67 ---RCQSGSGHGIASDWHL---EGSGG-----NDA----- 91
Db 1077 QDNCKNGAHTDAVNGYTCICPEGYSGLFCEFPMPVLPRTSPCDNFDQNGAQCIVRI 1136
QY 92 -----PGQYGA-----YFHDGGLAFPGHVFGRSLPEVPET-IELEVRTSTAS 133
Db 1137 NEPICQCLPGYQGEKCEKLVSNFINKESYLQIP-----SAKVRPQTWITLQIATDSDS 1190
QY 134 GLLHMQGVEGEAQGKDFISLGLQDHLVFRYQLSGEARLV-SEDPINDGEWHVITAL 192
Db 1191 GILLYKG-----DKDHIARELYRGRVRSYDTGSHPASAIYSVETINDGNPHIVELL 1242
QY 193 REGRGSIQVDG--BELVSGRSPGNVAVNAKGSVYIGAP---DVATL--TGGRFSSGI 245
Db 1243 ALDQSLSLVDGNGPKIITNLS--KQSTLNFDSPLYVGMFGKSNVSLRQAPQNGTSF 1300
QY 246 TGCVRNVLVHS 256
Db 1301 HGCIRNLYINS 1311

RESULT 14
US-09-540-153-2
; Sequence 2, Application US/09540153
; Patent No. 6270995
; GENERAL INFORMATION:
```

```
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,153
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/191,647
; PRIOR FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: human
US-09-540-153-2

Query Match 16.2%; Score 253.5; DB 3; Length 1525;
Best Local Similarity 24.8%; Pred. No. 1.6e-14;
Matches 92; Conservative 48; Mismatches 100; Indels 131; Gaps 18;

QY 1 CERPCQCHGATC-MPAGEYE-FQCLCRDGFPGDLCE-----HEENPC----- 40
Db 957 CISNPKCHGGTCHLKEGEDGFWCICADGFECEVNVDDCEDNCCNNSTCVDGINNY 1016
QY 41 -----QLREPCCLHGTC-----QGTCLCLPGFSGP----- 66
Db 1017 TCLCPPEYTGELCEKLDLFCQDLNFCQHDSKILTPKFKDCCTPGVYGEHCHDIDFDC 1076
QY 67 ---RCQSGSGHGIASDWHL---EGSGG-----NDA----- 91
Db 1077 QDNCKNGAHTDAVNGYTCICPEGYSGLFCEFPMPVLPRTSPCDNFDQNGAQCIVRI 1136
QY 92 -----PGQYGA-----YFHDGGLAFPGHVFGRSLPEVPET-IELEVRTSTAS 133
Db 1137 NEPICQCLPGYQGEKCEKLVSNFINKESYLQIP-----SAKVRPQTWITLQIATDSDS 1190
QY 134 GLLHMQGVEGEAQGKDFISLGLQDHLVFRYQLSGEARLV-SEDPINDGEWHVITAL 192
Db 1191 GILLYKG-----DKDHIARELYRGRVRSYDTGSHPASAIYSVETINDGNPHIVELL 1242
QY 193 REGRGSIQVDG--BELVSGRSPGNVAVNAKGSVYIGAP---DVATL--TGGRFSSGI 245
Db 1243 ALDQSLSLVDGNGPKIITNLS--KQSTLNFDSPLYVGMFGKSNVSLRQAPQNGTSF 1300
QY 246 TGCVRNVLVHS 256
Db 1301 HGCIRNLYINS 1311

RESULT 15
US-09-191-647-9
; Sequence 9, Application US/09191647
; Patent No. 6046015
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/191,647
; CURRENT FILING DATE: 1998-11-13
; EARLIER APPLICATION NUMBER: 60/065,544
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: 60/081,057
; EARLIER FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
```

```

; LENGTH: 735
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-191-647-9

Query Match      15.9%; Score 249; DB 3; Length 735;
Best Local Similarity 26.7%; Pred. No. 1.6e-14;
Matches 84; Conservative 48; Mismatches 115; Indels 68; Gaps 17;

QY 5 PCOHGATCMA-GEYEFQCLCRDGFGLCEHENPCOLREPCLHGGTC---QGTRCLC 59
Db 320 PCENNGKCIPINGSY--SCMSPGFTGNCTNIDDCNVE-CQNGSCVDGILSYDCLC 376
QY 60 LPGFSG-----PRCQ---GSGHGIA---ESDWHL---EGSGGNDAPGQ 94
Db 377 RFGYAGQYCEIPEPMMDMEYQKTDACQSQACQGCVCASQNSDFTCKCHEGFGPSCDRQ 436
QY 95 YGAYFHDDG-FLAFPGHVFSRSLPEVPE-TIELEVRTSTASGLLLWQGVVEGEAGQKDF 152
Db 437 MSVGFKNFCAYLALD-----FLASDGTITWLTTSKIGILLYYGD-----HF 480
QY 153 ISLGLQDGHVPRYQIGSGEA-RLVSEDPINDGEHVRVTLREGRRSIQVDGE--ELVS 209
Db 481 VSAELYDGRVKLVYIGNFPASHMYSSVKVNDGLFHRISIRTSERKCFQLQIDKNFVQIVE 540
QY 210 GRSPGPNVAVNAKGSVYIGGAP-----DVATLTGGRFSSGITGVKNLVLSARPAGPPP 264
Db 541 NSGKSDQLITKGMELYIGGLEPIEKSDAKRRFHVKNSESLKGCISITINEV----- 593
QY 265 QFLDLQHRAQAGANT 279
Db 594 -PINLQ-QALENVNT 606

```

Search completed: March 9, 2004, 17:23:33
Job time: 12.1326 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:11:22 ; Search time 34.5822 Seconds
(without alignments)
2320.373 Million cell updates/sec

Title: US-10-006-011a-9

Perfect score: 1566

Sequence: 1 CEEQPCQHGATCMPAGEYEF.....QPLDLQHRQAQANTRPCPS 284

Scoring table: ELOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseq1980s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001s.*
- 5: Geneseq2002s.*
- 6: Geneseq2003as.*
- 7: Geneseq2003bs.*
- 8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1566	100.0	4391	6	AAE34390	Human per	
2	1559	99.6	4393	4	AAAB31889	Amino aci	
3	1543	98.5	4436	4	ABG23265	Novel hum	
4	1030	65.8	195	4	AAAB31890	Amino aci	
5	357.5	22.8	1931	6	ABU52400	Human GPC	
6	343	21.9	238	4	AAU18145	Novel hum	
7	343	21.9	238	4	AAU17011	Human nov	
8	343	21.9	238	4	ABBI10450	Human nov	
9	343	21.9	238	4	AAU19961	Novel hum	
10	343	21.9	238	5	ABJ05772	Novel hum	
11	343	21.9	238	5	ABP67037	Human pol	
12	343	21.9	375	4	AAU07421	Novel hum	
13	343	21.9	432	4	AAU18102	Novel hum	
14	343	21.9	432	4	AAU18338	Human nov	
15	343	21.9	432	4	ABBI10233	Human nov	
16	343	21.9	432	4	AAU19905	Novel hum	
17	343	21.9	432	5	ABJ05729	Novel hum	
18	343	21.9	432	5	ABP66820	Human pol	
19	343	21.9	463	7	ABD64904	Human pro	
20	343	21.9	671	6	ABP58231	Human cel	
21	343	21.9	1009	7	ADE28105	Human NTR	
22	340.5	21.7	406	5	ABP41801	Human ova	
23	338.5	21.6	2053	7	ADC39154	Novel hum	
24	338.5	21.6	2143	7	ADC39164	Novel hum	
25	334	21.3	416	2	AAJ73993	Human pro	

26	334	21.3	1565	7	ADC39156	Novel hum
27	334	21.3	1566	7	ADC39166	Novel hum
28	330	21.1	492	2	AAW26609	Human agr
29	330	21.1	819	5	ABB72291	Rat prote
30	310	19.8	1544	7	ADC39162	Novel hum
31	303	19.3	1741	5	ABF43859	Human men
32	292	18.6	1298	4	ABB61948	Drosophil
33	292	18.6	1508	2	AAJ27141	Human sli
34	292	18.6	1508	2	AAW96706	Protein s
35	292	18.6	1508	2	AAJ04138	Human sli
36	292	18.6	1534	2	AAW46966	Amino aci
37	292	18.6	1534	2	AAJ27144	Human sli
38	292	18.6	1534	2	AAW96707	Protein s
39	292	18.6	1534	2	AAJ04139	Human sli
40	292	18.6	1534	7	ADE54473	Human pro
41	290.5	18.6	210	4	AAU19782	Human nov
42	290.5	18.6	210	5	ABP48002	Human pol
43	290.5	18.6	210	7	ADC10964	Human ext
44	281	17.9	4072	4	ABB63614	Drosophil
45	279.5	17.8	152	5	ABP56528	BM hepara

ALIGNMENTS

RESULT 1
AAE34390
ID AAE34390 standard; protein; 4391 AA.

AC AAE34390;

XX 14-MAY-2003 (first entry)

DT Human perlecan protein.

DE Human perlecan protein.

XX Human; diagnosis; osteoarthritis; rheumatoid arthritis; perlecan.

XX Homo sapiens.

XX WO200295415-A2.

XX 28-NOV-2002.

XX 22-MAY-2002; 2002WO-EP005612.

XX 23-MAY-2001; 2001GB-00012626.

XX (OSTE-) OSTROMETER BIO TECH AS.

XX Christgau S, Henriksen DB, Cloos PAC;

XX WPI; 2003-140389/13.

XX An assay for the diagnosis or assessment of the severity of osteoarthritis or rheumatoid arthritis comprising detecting an isomerized or optically inverted protein in a sample.

XX Disclosure; Page 46-67; 106pp; English.

XX The invention relates to an assay for the diagnosis or assessment of the severity of osteoarthritis or rheumatoid arthritis. The assay involves measuring (in a biological sample) the amount or presence of an isomerized or optically inverted protein or one or more isomerized or optically inverted fragments from proteins such as perlecan, biglycan, decorin, fibrillin-1 or procadherin. The assay is useful for the diagnosis or assessment of the severity of osteoarthritis or rheumatoid arthritis. The present sequence is human perlecan protein.

XX Sequence 4391 AA;

Query Match 100.0%; Score 1566; DB 6; Length 4391;

Best Local Similarity 100.0%; Pred. No. 8.2e-110; Indels 0; Gaps 0;
Matches 284; Conservative 0; Mismatches 0;

QY 1 CEROPCHGATCWPAGEYEFQCLCRDGFKGDLCHEENPCQLREPCQLHGCTCGTRCLCL 60
 DB 4108 CEROPCHGATCWPAGEYEFQCLCRDGFKGDLCHEENPCQLREPCQLHGCTCGTRCLCL 4167
 QY 61 PGSGPRCCQSGHGIAESDWHLESGGNDAPQGYGAYFHDDGFLAPPGHVFPSLSPEVP 120
 DB 4168 PGSGPRCCQSGHGIAESDWHLESGGNDAPQGYGAYFHDDGFLAPPGHVFPSLSPEVP 4227
 QY 121 ETIELEVRTSTASGLLLWQGVGEAGQKDFISLGLQDGHLYFRYQLGSGEARLVSEDP 180
 DB 4228 ETIELEVRTSTASGLLLWQGVGEAGQKDFISLGLQDGHLYFRYQLGSGEARLVSEDP 4287
 QY 181 INDGEHVRVTLALREGRRGSIQVDGEELVSGRSPGNVAVNAKGSYVIGGAPDVATLTGGR 240
 DB 4288 INDGEHVRVTLALREGRRGSIQVDGEELVSGRSPGNVAVNAKGSYVIGGAPDVATLTGGR 4347
 QY 241 FSSGITGCVKNLVLSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
 DB 4348 FSSGITGCVKNLVLSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391

RESULT 2

AA31889 standard; protein; 4393 AA.

AA31889;

15-MAY-2001 (first entry)

Amino acid sequence of a human protein.

Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 rheumatoid polyarthritis; lupus erythematosus; gene therapy.

Homo sapiens.

WO200105422-A2.

25-JAN-2001.

17-JUL-2000; 2000WO-FR002057.

15-JUL-1999; 99FR-00009372.

(INMR) BIOMERIEUX STELHYS.

Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

WPI; 2001-159475/16.

Detecting, preventing and treating degenerative, neurological and
 autoimmune diseases, particularly multiple sclerosis, using specified
 polypeptides or related nucleic acid or ligand.

Claim 1; Page 138-152; 209pp; French.

The present sequence represents a human protein, which is used in the
 method of the invention. The specification describes a method which uses
 at least one polypeptide or polynucleotide sequence belonging to the
 perlecan, precursor of the retinol-binding plasma protein, precursor of
 the ganglioside GM2 activator, calgranulin B or saposin B protein
 families. The method is used for detecting, preventing or treating a
 degenerative, neurological and/or auto-immune disease. The
 polynucleotides and polypeptides are used for diagnosis, prognosis,
 prevention and treatment of multiple sclerosis (in its various forms and
 phases). They may also be useful in cases of e.g. Alzheimer's and
 Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 polyarthritis and lupus erythematosus, including use as vaccines and in
 gene therapy (expression of sense or antisense sequences). They can also

CC be used to assess efficacy of potential therapeutic agents, particularly
 compounds that reduce or inhibit toxicity towards glial cells

Query Match 99.6%; Score 1559; DB 4; Length 4393;

Best Local Similarity 99.3%; Pred. No. 2.8e-109;

Matches 282; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Sequence 4393 AA;

QY 1 CEROPCHGATCWPAGEYEFQCLCRDGFKGDLCHEENPCQLREPCQLHGCTCGTRCLCL 60

DB 4110 CEROPCHGATCWPAGEYEFQCLCRDGFKGDLCHEENPCQLREPCQLHGCTCGTRCLCL 4169

QY 61 PGSGPRCCQSGHGIAESDWHLESGGNDAPQGYGAYFHDDGFLAPPGHVFPSLSPEVP 120

DB 4170 PGSGPRCCQSGHGIAESDWHLESGGNDAPQGYGAYFHDDGFLAPPGHVFPSLSPEVP 4229

QY 121 ETIELEVRTSTASGLLLWQGVGEAGQKDFISLGLQDGHLYFRYQLGSGEARLVSEDP 180

DB 4230 ETIELEVRTSTASGLLLWQGVGEAGQKDFISLGLQDGHLYFRYQLGSGEARLVSEDP 4289

QY 181 INDGEHVRVTLALREGRRGSIQVDGEELVSGRSPGNVAVNAKGSYVIGGAPDVATLTGGR 240

DB 4290 INDGEHVRVTLALREGRRGSIQVDGEELVSGRSPGNVAVNAKGSYVIGGAPDVATLTGGR 4349

QY 241 FSSGITGCVKNLVLSARPGAPPPQPLDLQHRAQAGANTRPCPS 284

DB 4350 FSSGITGCVKNLVLSARPGAPPPQPLDLQHRAQAGANTRPCPS 4393

RESULT 3

ABG23265 standard; protein; 4436 AA.

ABG23265;

18-FEB-2002 (first entry)

Novel human diagnostic protein #23256.

Human; chromosome mapping; gene mapping; gene therapy; forensic;

Food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US008631.

31-MAR-2000; 2000US-00540217.

23-AUG-2000; 2000US-00649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

N-PSDB; AAS87452.

New isolated polynucleotide and encoded polypeptides, useful in

diagnostics, forensics, gene mapping, identification of mutations

responsible for genetic disorders or other traits and to assess

biodiversity.

Claim 20; SEQ ID NO 53624; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II)

sequences. (I) is useful as hybridisation probes, polymerase chain

reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

and in recombinant production of (II). The polynucleotides are also used

in diagnostics as expressed sequence tags for identifying expressed

CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activities. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC patent did not appear in the invention. Note: The sequence data for this
 CC electronic format directly from WFO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 4436 AA;

Query Match 98.5%; Score 1543; DB 4; Length 4436;
 Best Local Similarity 96.2%; Pred. No. 4.7e-108;
 Matches 282; Conservative 0; Mismatches 1; Indels 10; Gaps 1;
 QY 1 CERQPCQHGATCMAPAGEYEFQCLRDGFKGDLCEHENPCQLRPPCLHGTCQGTCLCL 60
 DB 4143 CERQPCQHGATCMAPAGEYEFQCLRDGFKGDLCEHENPCQLRPPCLHGTCQGTCLCL 4202
 QY 61 PGFSGPRCOQSGHGIAESDWHLESGGNDAPGQYGFYHDDGFLAFPGHVFGRSLPEVP 120
 DB 4203 PGFSGPRCOQSGHGIAESDWHLESGGNDAPGQYGFYHDDGFLAFPGHVFGRSLPEVP 4262
 QY 121 ETIELEVTSTASGLLLWQGVVEGAGQKDFISLGLQDGHVLF-----RYQLGS 170
 DB 4263 ETIELEVTSTASGLLLWQGVVEGAGQKDFISLGLQDGHVLF-----RYQLGS 4322
 QY 171 GEARLVSEDPINDGEHVRVTALREGRGSIQVDGEELVSGSPGNVAVNAKGSVYIGGA 230
 DB 4323 GEARLVSEDPINDGEHVRVTALREGRGSIQVDGEELVSGSPGNVAVNAKGSVYIGGA 4382
 QY 231 PVAVTLTGRFSSGTTGCVKNLVLSARPGAPPPQDLDQHRAQAGANTRP 283
 DB 4383 PDVATLTGRFSSGTTGCVKNLVLSARPGAPPPQDLDQHRAQAGANTRP 4435

RESULT 4
 AAB31890
 ID AAB31890 standard; protein; 195 AA.
 XX
 AC AAB31890;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of the C-terminal of the human perlecan protein.
 XX
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200105422-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 17-JUL-2000; 2000WO-FR002057.
 XX
 PR 15-JUL-1999; 99FR-00009372.
 XX
 PA (INVR) BIOMERIEUX STELHYS.
 XX
 PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX WPI; 2001-159475/16.
 DR N-PSDB; AAF54728.
 XX
 PT Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 PS Claim 1; Page 152-153; 209pp; French.
 XX
 CC The present sequence represents a human polypeptide, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 195 AA;

Query Match 65.8%; Score 1030; DB 4; Length 195;
 Best Local Similarity 100.0%; Pred. No. 1.5e-70;
 Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 90 DAPGQYGFYHDDGFLAFPGHVFGRSLPEVPETIEVLTSTASGLLLWQGVVEGAGQ 149
 DB 1 DAPGQYGFYHDDGFLAFPGHVFGRSLPEVPETIEVLTSTASGLLLWQGVVEGAGQ 60
 QY 150 KDFISLGLQDGHVLYRQVLSGEARLVSEDPINDGEHVRVTALREGRGSIQVDGEELVS 209
 DB 61 KDFISLGLQDGHVLYRQVLSGEARLVSEDPINDGEHVRVTALREGRGSIQVDGEELVS 120
 QY 210 GRSPGNVAVNAKGSVYIGGAPDVATLTGGRFSSGTTGCVKNLVLSARPGAPPPQDLD 269
 DB 121 GRSPGNVAVNAKGSVYIGGAPDVATLTGGRFSSGTTGCVKNLVLSARPGAPPPQDLD 180
 QY 270 QHRAQAGANTRP 284
 DB 181 QHRAQAGANTRP 195

RESULT 5
 ABUS2400
 ID ABUS2400 standard; protein; 1931 AA.
 XX
 AC ABUS2400;
 XX
 DT 03-MAR-2003 (first entry)
 XX
 DE Human GPCR related protein NOV40a.
 XX
 KW Human; NOVX; G-protein coupled receptor; GPCR; cancer; cytostatic.
 XX
 OS Homo sapiens.
 XX
 PN WO200279398-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 08-MAR-2002; 2002WO-US007355.
 XX
 PR 08-MAR-2001; 2001US-0274194P.
 PR 08-MAR-2001; 2001US-0274281P.
 PR 08-MAR-2001; 2001US-0274322P.
 PR 09-MAR-2001; 2001US-0274849P.

13-MAR-2001; 2001US-0275578P.
 13-MAR-2001; 2001US-0275579P.
 13-MAR-2001; 2001US-0275601P.
 14-MAR-2001; 2001US-0276000P.
 16-MAR-2001; 2001US-0276776P.
 19-MAR-2001; 2001US-0276894P.
 20-MAR-2001; 2001US-0277239P.
 20-MAR-2001; 2001US-0277327P.
 20-MAR-2001; 2001US-0277338P.
 21-MAR-2001; 2001US-0277791P.
 22-MAR-2001; 2001US-0277833P.
 23-MAR-2001; 2001US-0278152P.
 26-MAR-2001; 2001US-0278894P.
 27-MAR-2001; 2001US-0278999P.
 27-MAR-2001; 2001US-0279036P.
 30-MAR-2001; 2001US-0280233P.
 02-APR-2001; 2001US-0280802P.
 02-MAY-2001; 2001US-0288052P.
 02-MAY-2001; 2001US-0288066P.
 02-MAY-2001; 2001US-0288228P.
 17-MAY-2001; 2001US-0291766P.
 07-JUN-2001; 2001US-0296693P.
 08-JUN-2001; 2001US-0296856P.
 05-JUL-2001; 2001US-0303230P.
 05-JUL-2001; 2001US-0303377P.
 08-AUG-2001; 2001US-0310913P.
 13-AUG-2001; 2001US-0311978P.
 14-AUG-2001; 2001US-0312191P.
 16-AUG-2001; 2001US-0312916P.
 17-AUG-2001; 2001US-0313182P.
 20-AUG-2001; 2001US-0313626P.
 21-AUG-2001; 2001US-0314018P.
 21-AUG-2001; 2001US-0315227P.
 10-SEP-2001; 2001US-0318403P.
 10-SEP-2001; 2001US-0318510P.
 14-SEP-2001; 2001US-0322396P.
 14-SEP-2001; 2001US-0322396P.
 27-SEP-2001; 2001US-0325378P.
 09-NOV-2001; 2001US-0332486P.
 09-NOV-2001; 2001US-0345399P.
 07-MAR-2002; 2002US-00094866.
 (CURA-) CURAGEN CORP.
 Kekuda R, Tchernev VT, Liu X, Spytek KA, Patturajan M, Burgess CE, Vernet CM, Li L, Gorman L, Malyankar UM, Boldog PL; Guo X, Shenoy SG, Radigan M, Taupier RJ, Miller CE, Casman SJ; Pena CE, Gangolli EA, Gusev V, Smithson G, Zerhusen BD, Gerlach V; Pochart PF, Fernandes ER, Shimkets RA, Rastelli L, Spaderna SK; Larochelle WJ, Zhong M, Khramtsov NV, Voss EZ, Herrmann JL; WPI; 2003-058423/05.
 N-PSDB; ABX70491.
 NOVX polypeptides and polynucleotides, useful for treating a syndrome related to a human disease associated with the NOVX polypeptide e.g., cancer.
 Claim 1; Page 223; 413pp; English.
 The present invention relates to the isolation of novel human polypeptides referred to as NOVX (NOV1-NOV44), variants of these proteins, and the polynucleotide sequences encoding them. The NOVX proteins of the invention are G-protein coupled receptor (GPCR) related proteins. The sequences of the invention are useful in the manufacture of a medicament for treating a syndrome related to a human disease associated with the polypeptides e.g. cancer. ABUS2311-ABUS2408 represent the human NOVX proteins of the invention
 Sequence 1931 AA;
 Query Match 22.8%; Score 357.5; DB 6; Length 1931;
 Best Local Similarity 35.4%; Pred. No. 3.1e-18;

Matches 91; Conservative 29; Mismatches 110; Indels 27; Gaps 8;
 QY 1 CEROPCOHGATCMPEAGEVEYFCCLCRDGFKGDLCHEHENPCQLRPGTCLHGTCQ-----GT 55
 Db 1439 CLPNECHGAPQCNLEAGRFHCPCQPPGVRGPTCADEKSPCQ-FNPCHGAAPCRVLPBGGA 1497
 QY 56 RCLCLPGSPGRCQGGSHGTAESDWELESGGNDAPGQYGAFFHDDGFLAPG-HVFSR 114
 Db 1498 QCECPGREGTFCQTAS-----GQDGGPFLADFNFGFSHLELGLHTFAR 1542
 QY 115 SLPEVPETIELEVTSTASGLLLMQGVEGAGGQKDFISLGLQDGHVFRYQLGSGEAR 174
 Db 1543 DLGE-KMALEVVVFLARGPSGLLYNGQKTD--GKG-DFVSLALDRRLRLEFRYDLGKGA 1598
 QY 175 LVSEDPINDGEMHVTALREGRGSIQVDGEELVSGSPGNVAVNAKGSVITGGAPDVA 234
 Db 1599 IRSREFVTGATRVSLERNRGKALRGDGPVRLGESPVFHVVLNKLKEPLYVGGAPDFS 1658
 QY 235 TLT-GGRFSSGITGVK 250
 Db 1659 KLARAAVSSGFDGAIQ 1675
 RESULT 6
 AAU18145
 ID AAU18145 standard; protein; 238 AA.
 XX AC AAU18145;
 XX DT 21-NOV-2001 (first entry)
 XX DE Novel human uterine motility-association polypeptide #52.
 XX KW Human; uterine motility-association disorder; uterus; pregnancy; labour;
 XX KW menstrual cycle; gene therapy.
 XX OS Homo sapiens.
 XX PN WO200155201-A1.
 XX PD 02-AUG-2001.
 XX PF 17-JAN-2001; 2001WO-US001317.
 XX PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0196123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 30-JUN-2000; 2000US-0214886P.
 PR 07-JUL-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 11-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 14-AUG-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.

57 21 20 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1

Query.Match 21.9%: Score 343: DB 4: Length 238:

Qy 146 AGCKDFISLIGDGHVFRYOLGSGEARLVSEDDINDGEWHRVTALEGRGSIQVDGE 205
 Db 106 MRNSDFISLIGDGHVFRYOLGSGEARLVSEDDINDGEWHRVTALEGRGSIQVDGE 165
 Qy 206 ELVSGSPGPNVAVNAKGVYIGGAPDVATLTGGRFSSGTCVKNLVHSAREFGAPPPQ 265
 Db 166 GARTKSPGMRLQNLGALYVGMKEIALHTNRQYMRGLVGCISHFTLST-----DY 218
 Qy 266 PLQLQHRQAQAGANTREC 282
 Db 219 HISLVEDAVDGKKNITC 235

RESULT 7
 AAU17011
 ID AAU17011 standard; protein; 238 AA.

AC AAU17011;
 DT 07-NOV-2001 (first entry)
 XX Human novel secreted protein, SEQ ID 252.
 DE Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
 KW Cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective;
 KW antibacterial; virucide; fungicide; ophthalmological; vulnary;
 KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 KW cerebral ischaemia; angiogenesis; nervous system disorder;
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; epithelial cell proliferation; skin ageing; food additive;
 KW preservative; antiproliferative.

OS Homo sapiens.
 XX WO200155441-A2.
 PX 02-AUG-2001.

17-JAN-2001; 2001WO-US0001320.

31-JAN-2000; 2000US-0179065P.
 04-FEB-2000; 2000US-0180628P.
 24-FEB-2000; 2000US-0184654P.
 02-MAR-2000; 2000US-0186350P.
 16-MAR-2000; 2000US-0189874P.
 17-MAR-2000; 2000US-0190076P.
 18-APR-2000; 2000US-0198123P.
 19-MAY-2000; 2000US-0205515P.
 07-JUN-2000; 2000US-0209467P.
 28-JUN-2000; 2000US-0214886P.
 30-JUN-2000; 2000US-0215135P.
 07-JUL-2000; 2000US-0216847P.
 07-JUL-2000; 2000US-0216880P.
 11-JUL-2000; 2000US-0217487P.
 11-JUL-2000; 2000US-0217496P.
 14-JUL-2000; 2000US-0218290P.
 26-JUL-2000; 2000US-0220963P.
 26-JUL-2000; 2000US-0220964P.
 14-AUG-2000; 2000US-0224518P.
 14-AUG-2000; 2000US-0225113P.
 14-AUG-2000; 2000US-0225213P.
 14-AUG-2000; 2000US-0225266P.
 14-AUG-2000; 2000US-0225268P.
 14-AUG-2000; 2000US-0225270P.
 14-AUG-2000; 2000US-0225447P.
 14-AUG-2000; 2000US-0225757P.
 14-AUG-2000; 2000US-0225758P.
 14-AUG-2000; 2000US-0225759P.
 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.

PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 05-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 06-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231113P.
 PR 08-SEP-2000; 2000US-0231141P.
 PR 08-SEP-2000; 2000US-0232089P.
 PR 12-SEP-2000; 2000US-0232081P.
 PR 14-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 21-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234397P.
 PR 25-SEP-2000; 2000US-0234398P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236357P.
 PR 29-SEP-2000; 2000US-0236358P.
 PR 29-SEP-2000; 2000US-0236359P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.

PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246612P.
 PR 17-NOV-2000; 2000US-0248207P.
 PR 17-NOV-2000; 2000US-0248208P.
 PR 17-NOV-2000; 2000US-0248209P.
 PR 17-NOV-2000; 2000US-0248210P.
 PR 17-NOV-2000; 2000US-0248211P.
 PR 17-NOV-2000; 2000US-0248212P.
 PR 17-NOV-2000; 2000US-0248213P.
 PR 17-NOV-2000; 2000US-0248214P.
 PR 17-NOV-2000; 2000US-0248215P.
 PR 17-NOV-2000; 2000US-0248216P.
 PR 17-NOV-2000; 2000US-0248217P.
 PR 17-NOV-2000; 2000US-0248218P.
 PR 17-NOV-2000; 2000US-0248219P.
 PR 17-NOV-2000; 2000US-0248244P.
 PR 17-NOV-2000; 2000US-0248245P.
 PR 17-NOV-2000; 2000US-0248246P.
 PR 17-NOV-2000; 2000US-0248265P.
 PR 17-NOV-2000; 2000US-0248297P.
 PR 17-NOV-2000; 2000US-0248299P.
 PR 01-DEC-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-476161/51.
 DR N-PSDB; ABA06672.
 XX
 PT Isolated nucleic acid molecule encoding an inflammation-associated
 PT polypeptide is used in preventing, treating or ameliorating a medical
 PT condition.
 XX
 PS Claim 11; SEQ ID NO 758; 859pp + Sequence Listing, English.
 XX
 CC The present invention provides human cDNAs, proteins and related genomic
 CC DNAs. These can be used in the treatment of neural, immune system,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal and proliferative disorders and inflammation. The present sequence
 CC is a protein of the invention
 XX
 SQ Sequence 238 AA;
 Query Match 21.9%; Score 143; DB 4; Length 238;
 Best Local Similarity 33.5%; Pred. No. 3.6e-18;
 Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
 Qy 34 EHEHPCQLRPFCLHGGTC---GGTCLCLPFGSGPRCQSGHGIABSDWHLGSGGN 89
 Db 11 ENAHPC-VRAPCAHGGSCRPKEGYDCPLGFLHCOKAITEA-----56
 Qy 90 DAP---GQYAYFDDGFLAFPGHVFPSRLPEVPTIELEVTSTASGLLLWQVEUGE 145
 Db 57 EIPQFGRSYLTNDNPDLKEVSG---SR-----NVFMFXTAKDGLLWRG--DSP 105
 Qy 146 ACQGDQFISLGLDGHVFRYQLSGEARLVSDPINDGEWHVYALREGRRSIQDTGE 205
 Db 106 MRPNSDFISLGLRDGALVFSYNLGSGVASIMWNGSFNDGWRHVRKAVRDGQSKITVDDY 165

QY 206 ELVSGRPGPNVAVNAKSGVYIGGAPDVATLTGRFSSGTCVKNLVHSARPGAPPPQ 265
 Db 166 GARTGKSPGNMRLQNLINGALYVGGMKEIALHTNRQVMRGLVGCISHFTLST-----DY 218
 QY 266 PLDLOHRAQAQAGANTRPC 282
 Db 219 HISLVEDAVDGKINTC 235

RESULT 9
 ID AAU19961 standard; protein; 238 AA.

AC AAU19961;
 XX
 XX
 XX 06-DEC-2001 (first entry)
 DE Novel human calcium-binding protein #70.
 DE Human; calcium-binding protein; calcium flux; neurological disease;
 KW immune dysfunction; digestive disorder; neoplastic disease;
 KW blood disorder; infectious disease; gene therapy; immunosuppressive;
 KW antiarthritic; cytostatic; vasotropic; antibacterial; nontropic;
 KW virucide.
 XX
 OS Homo sapiens.
 XX
 XX
 XX WO200155304-A2.
 XX
 XX
 PD 02-AUG-2001.
 XX
 XX 17-JAN-2001; 2001WO-US0001302.

PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 11-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 14-JUL-2000; 2000US-0217496P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225367P.
 PR 14-AUG-2000; 2000US-0225368P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225476P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226688P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-022824P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 12-SEP-2000; 2000US-0232081P.
 PR 14-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 13-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.

34	QY	EHEENPCQLRPPCLHGTC	-----GCTCTCLPQSPGPRCOOGSCHCIAESDWHLSGSGN	89
		: : : : : : :	: : : : : : :	
11	Db	ENAAHP--VRAPCAHGGSCRPRKGYDCDPLFEGULHCOAIEAI	-----	56
		: : : : : : :	: : : : : : :	
90	QY	DAP-----QQYGAIFYHDGFLAPPGHVSRSFLPEVETIEVETSTASGLLLMQGVVE	145	
		: : : : : : :	: : : : : : :	
57	Db	EIEPQIGRSYLTYNPDILKVSQ-----NVFMRFKTKADGLLWRG--DSP	105	
		: : : : : : :	: : : : : : :	
146	QY	ACQGGKDFISLGLQDGLHLYFRYQLQSGSEARLVSEDPINDGSEHWRTALRSRGRGSIQVDGE	205	

Db	106	WRPNSDFSLGLRDLGVFSYINLGSVASIWNVNGSFNDGRWHRVXAVRDGSGKITVDYD	166
QY	206	ELVSGRSPGNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSAAPGAPPPQ	265
Db	166	CARTGKSGMWRQLNINALYVGGMKETIALFTNRYMRGLVGCISHFTLST-----DY	218
QY	266	PLDLQHRACAGANTRPC	282
Db	219	HISLVEDAVDGKNITC	235
RESULT 11			
ABP67037			
ABP67037 standard; protein; 238 AA.			
XX	AC	ABP67037;	
XX	AC		
DT	09-DEC-2002	(first entry)	
XX			
XX	Human polypeptide SEQ ID NO 758.		
XX	Human;	neotropic; neuroprotective; cytostatic; dermatological; virucide;	
XX	immunosuppressive;	antiinflammatory; anti-HIV; antibacterial; vulnerary;	
KW	antiparkinsonian;	antickling; antianaemic; antiarthritic; cancer;	
KW	antirheumatic;	hepatotropic; cerebrotective; antiinflammatory;	
KW	antiallergic;	acidulabec; antituler; anticonvulsant; antifungal;	
KW	antiparasitic;	cardiant; immune disorder; cardiovascular disorder;	
KW	neurological disease;	infection; nephrotropic; gene therapy; vaccine.	
XX			
XX	Homo sapiens.		
OS			
XX			
XX	US2002090672-A1.		
PN			
XX			
PD	11-JUL-2002.		
XX			
XX	17-JAN-2001;	2001US-00764853.	
XX			
31	JAN-2000;	2000US-0179065P.	
PR	04-FEB-2000;	2000US-0180628P.	
PR	28-JUN-2000;	2000US-0214896P.	
PR	07-JUL-2000;	2000US-0216647P.	
PR	07-JUL-2000;	2000US-0216880P.	
PR	11-JUL-2000;	2000US-0217487P.	
PR	11-JUL-2000;	2000US-0217496P.	
PR	14-JUL-2000;	2000US-0218290P.	
PR	26-JUL-2000;	2000US-0220963P.	
PR	14-AUG-2000;	2000US-0224518P.	
PR	14-AUG-2000;	2000US-0224519P.	
PR	14-AUG-2000;	2000US-0225267P.	
PR	14-AUG-2000;	2000US-0225288P.	
PR	14-AUG-2000;	2000US-0225270P.	
PR	14-AUG-2000;	2000US-0225447P.	
PR	14-AUG-2000;	2000US-0225757P.	
PR	14-AUG-2000;	2000US-0225758P.	
PR	22-AUG-2000;	2000US-0226868P.	
PR	30-AUG-2000;	2000US-0228924P.	
PR	01-SEP-2000;	2000US-0229287P.	
PR	01-SEP-2000;	2000US-0229343P.	
PR	01-SEP-2000;	2000US-0229344P.	
PR	01-SEP-2000;	2000US-0229345P.	
PR	05-SEP-2000;	2000US-0229509P.	
PR	05-SEP-2000;	2000US-0229513P.	
PR	08-SEP-2000;	2000US-0231413P.	
PR	21-SEP-2000;	2000US-0234223P.	
PR	21-SEP-2000;	2000US-0234274P.	
PR	25-SEP-2000;	2000US-0234997P.	
PR	27-SEP-2000;	2000US-0235834P.	
PR	29-SEP-2000;	2000US-0236327P.	
PR	29-SEP-2000;	2000US-0236367P.	
PR	29-SEP-2000;	2000US-0236368P.	
PR	29-SEP-2000;	2000US-0236369P.	

PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 13-OCT-2000; 2000US-0237040P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 XX
 PA (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX
 FI Rosen CA, Ruben SM, Barash SC;
 XX
 DR WPI; 2002-681727/73.
 DR N-PSDB; ABV84009.
 XX
 Novel polypeptide useful for diagnosis, prognosis, prevention, and treatment of immune, hyperproliferative, renal, respiratory, cardiovascular, reproductive, endocrine, gastrointestinal and neurological disorders.
 PT
 PT
 PT
 XX
 PS Claim 11; SEQ ID NO 758; 369pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABV83692-ABV84101) and proteins (ABP6710-ABP67129) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 CC
 SQ Sequence 238 AA;
 Query Match 21.9%; Score 343; DB 5; Length 238;
 Best Local Similarity 33.5%; Pred. No. 3.6e-18;
 Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
 QY 34 EHENPCQLREPCILGTC-----QVTRCLCLQFSQPCQOQSGHIAESDWHLEGGCN 89
 DB 11 ENAAHPC-VKAPCAHSGSGRKEGVDCCPLGFEGLHCQKAIIRAI-----56
 QY 90 DAP----GQYGAFFHDDGFLAFFGHVFSRLPEVETIEVRTSTASGLLLWQGVVEGE 145
 DB 57 EIPQIGRSYLTNDPILKRVSG--SRS-----NVFMRPKTKADGILLWRG--DSP 105
 QY 146 AGQCKDFISLGLDGLHVPYQYLGSGEARLVSEDPINDGEMHRTALREGRGSIQVDGE 205
 DB 106 MRPNDFISLGLDGLHVPYQYLGSGEARLVSEDPINDGEMHRTALREGRGSIQVDGE 165
 QY 206 ELVSGRFGPNVANKGSIYIGGAPDVAFLTGGRPSSITGCVKRLVHLSARPGAPPPQ 265
 DB 166 GARTGKSGFQWRLNINQALYVSGMKEIALHTNRQYMRGLVGCISHFTLST-----DY 218
 QY 266 PLDLQHQRAQAQANTPC 282

Db 219 HISLVEDAVDGNINTC 235
 RESULT 12
 AAU07421
 ID AAU07421 standard; protein; 375 AA.
 XX
 AC AAU07421;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human extracellular matrix (ECM) protein #4.
 XX
 KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic; antiproliferative; cardiant; vasotropic; cerebroprotective; nootropic; KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological; KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis; KW hyperproliferative disorder; neoplasm; cardiovascular disorder; KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection; KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn; KW wound healing; food additive.
 XX
 OS Homo sapiens.
 XX
 PN WO200179253-A1.
 XX
 PD 25-OCT-2001.
 XX
 PF 11-APR-2001; 2001WO-US011643.
 XX
 PR 18-APR-2000; 2000US-0198123P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Fiscella M, Shi Y, Ebner R, Ruben SM;
 DR WPI; 2001-611720/70.
 DR N-PSDB; AAS13846.
 XX
 PT New nucleic acids encoding extracellular matrix polypeptides, for PT diagnosing, treating, preventing or ameliorating human disorders and PT disease, such as, autoimmune, hyperproliferative or cardiovascular PT disorders.
 XX
 PS Claim 1; Page 297-298; 308pp; English.
 CC The invention relates to novel isolated polynucleotides (1) encoding CC extracellular matrix (ECM) polypeptides. (2) and a polypeptide encoded by CC (1) are used to prevent, treat or ameliorate a medical condition in e.g. CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They CC are also used in diagnosing a pathological condition or susceptibility to CC a pathological condition. The antibodies to the polypeptides can also be CC used in alleviating symptoms associated with the disorders and in CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated CC include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, CC angioneurosis, nervous system disorders e.g. Alzheimer's disease, CC infections caused by bacteria, viruses and fungi and ocular disorders CC e.g. corneal infection. The polypeptides can also be used to aid wound CC healing and epithelial cell proliferation, to prevent skin aging due to CC sunburn, to maintain organs before transplantation, for supporting cell CC culture of primary tissues, to regenerate tissues and in chemotaxis. The CC polypeptides can also be used as a food additive or preservative to CC increase or decrease storage capabilities. The present sequence CC represents the amino acid sequence of novel human extracellular matrix CC (ECM) protein #4
 XX
 SQ Sequence 375 AA;
 Query Match 21.9%; Score 343; DB 4; Length 375;

PR	18-AUG-2000;	2000US-02262709P;
PR	22-AUG-2000;	2000US-02266816P;
PR	22-AUG-2000;	2000US-02266861P;
PR	22-AUG-2000;	2000US-02271826P;
PR	23-AUG-2000;	2000US-02277009P;
PR	23-AUG-2000;	2000US-02282924P;
PR	30-SEP-2000;	2000US-02292878P;
PR	01-SEP-2000;	2000US-02293287P;
PR	01-SEP-2000;	2000US-02293434P;
PR	01-SEP-2000;	2000US-0229344P;
PR	01-SEP-2000;	2000US-0229345P;
PR	05-SEP-2000;	2000US-02295009P;
PR	05-SEP-2000;	2000US-02295133P;
PR	06-SEP-2000;	2000US-0230437P;
PR	06-SEP-2000;	2000US-0230439P;
PR	08-SEP-2000;	2000US-0231242P;
PR	08-SEP-2000;	2000US-0231243P;
PR	08-SEP-2000;	2000US-0231243P;
PR	08-SEP-2000;	2000US-0231413P;
PR	08-SEP-2000;	2000US-0231414P;
PR	08-SEP-2000;	2000US-0232080P;
PR	12-SEP-2000;	2000US-0232081P;
PR	12-SEP-2000;	2000US-0231968P;
PR	14-SEP-2000;	2000US-0232397P;
PR	14-SEP-2000;	2000US-0232398P;
PR	14-SEP-2000;	2000US-0232398P;
PR	14-SEP-2000;	2000US-0232400P;
PR	14-SEP-2000;	2000US-0232401P;
PR	14-SEP-2000;	2000US-0233063P;
PR	14-SEP-2000;	2000US-0233064P;
PR	14-SEP-2000;	2000US-0233065P;
PR	21-SEP-2000;	2000US-0234223P;
PR	21-SEP-2000;	2000US-0234274P;
PR	25-SEP-2000;	2000US-0234997P;
PR	25-SEP-2000;	2000US-0234998P;
PR	25-SEP-2000;	2000US-0234998P;
PR	26-SEP-2000;	2000US-0235482P;
PR	27-SEP-2000;	2000US-0235634P;
PR	27-SEP-2000;	2000US-0235634P;
PR	29-SEP-2000;	2000US-0236327P;
PR	29-SEP-2000;	2000US-0236327P;
PR	29-SEP-2000;	2000US-0236367P;
PR	29-SEP-2000;	2000US-0236368P;
PR	29-SEP-2000;	2000US-0236369P;
PR	29-SEP-2000;	2000US-0236370P;
PR	02-OCT-2000;	2000US-0236602P;
PR	02-OCT-2000;	2000US-0237037P;
PR	02-OCT-2000;	2000US-0237038P;
PR	02-OCT-2000;	2000US-0237039P;
PR	13-OCT-2000;	2000US-0237040P;
PR	13-OCT-2000;	2000US-0239935P;
PR	13-OCT-2000;	2000US-0239937P;
PR	20-OCT-2000;	2000US-0240960P;
PR	20-OCT-2000;	2000US-0241221P;
PR	20-OCT-2000;	2000US-0241785P;
PR	20-OCT-2000;	2000US-0241786P;
PR	20-OCT-2000;	2000US-0241787P;
PR	20-OCT-2000;	2000US-0241808P;
PR	20-OCT-2000;	2000US-0241809P;
PR	01-NOV-2000;	2000US-0241826P;
PR	01-NOV-2000;	2000US-0244617P;
PR	08-NOV-2000;	2000US-0246474P;
PR	08-NOV-2000;	2000US-0246475P;
PR	08-NOV-2000;	2000US-0246476P;
PR	08-NOV-2000;	2000US-0246478P;
PR	08-NOV-2000;	2000US-0246523P;
PR	08-NOV-2000;	2000US-0246524P;
PR	08-NOV-2000;	2000US-0246525P;
PR	08-NOV-2000;	2000US-0246526P;
PR	08-NOV-2000;	2000US-0246527P;
PR	08-NOV-2000;	2000US-0246528P;
PR	08-NOV-2000;	2000US-0246532P;
PR	08-NOV-2000;	2000US-0246533P;
PR	08-NOV-2000;	2000US-0246539P;
PR	08-NOV-2000;	2000US-02465610P;
PR	08-NOV-2000;	2000US-02465611P;

PR	08-NOV-2000;	2000US-0246613P.
PR	17-NOV-2000;	2000US-0249407P.
PR	17-NOV-2000;	2000US-0249408P.
PR	17-NOV-2000;	2000US-0249409P.
PR	17-NOV-2000;	2000US-0249410P.
PR	17-NOV-2000;	2000US-0249411P.
PR	17-NOV-2000;	2000US-0249412P.
PR	17-NOV-2000;	2000US-0249413P.
PR	17-NOV-2000;	2000US-0249414P.
PR	17-NOV-2000;	2000US-0249415P.
PR	17-NOV-2000;	2000US-0249416P.
PR	17-NOV-2000;	2000US-0249417P.
PR	17-NOV-2000;	2000US-0249418P.
PR	17-NOV-2000;	2000US-0249419P.
PR	17-NOV-2000;	2000US-0249424P.
PR	17-NOV-2000;	2000US-0249425P.
PR	17-NOV-2000;	2000US-0249426P.
PR	17-NOV-2000;	2000US-0249427P.
PR	17-NOV-2000;	2000US-0249429P.
PR	17-NOV-2000;	2000US-0249430P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251038P.
PR	05-DEC-2000;	2000US-0251198P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0256147P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0250497P.
PR	05-JAN-2001;	2001US-0254967B.
XX	(HUMA-) HUMAN GENOME SCI INC.	

Claim 11: SEO ID NO 179: 601pp: English:

Query Match 21.9%; Score 343; DB 4; Length 432;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:11:52 ; Search time 17.0517 Seconds
(without alignments)
2152.832 Million cell updates/sec

Title: US-10-006-011a-3

Perfect score: 3825

Sequence: 1 EIKITRPSADGMLYNGQ.....QPLDLQRAQAGANTRPCPS 705

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3825	100.0	4391	1	PGBM_HUMAN
2	3389.5	88.6	3707	1	PGBM_MOUSE
3	859	22.5	1959	1	AGRI_RAT
4	796.5	20.8	1355	1	AGRI_CHICK
5	667	17.4	1328	1	AGRI_DISOM
6	511	13.4	1375	1	UN52 CAEEL
7	441	11.5	3313	1	CLR3_RAT
8	421	11.0	1715	1	NX2A_RAT
9	420.5	11.0	1712	1	NX2A_HUMAN
10	412.5	10.8	1477	1	NX1A_HUMAN
11	405.5	10.6	1530	1	NX1A_BOVIN
12	403.5	10.5	1514	1	NX1A_RAT
13	403	10.5	3301	1	CLR3_MOUSE
14	397.5	10.4	1529	1	SLR2_HUMAN
15	397.5	10.4	3312	1	CLR2_HUMAN
16	396	10.4	1541	1	NX3A_HUMAN
17	394	10.3	1363	1	NX1A_CHICK
18	389	10.2	2144	1	CLR2_RAT
19	388	10.1	5147	1	FAT_DROME
20	386	10.1	5149	1	STAN_DROME
21	382	10.0	2923	1	CLR2_HUMAN
22	367	9.6	1578	1	NX3A_RAT
23	364	9.5	2920	1	CLR2_MOUSE
24	357	9.3	3110	1	LM2A_HUMAN
25	354.5	9.3	3110	1	CLR4_MOUSE
26	353.5	9.2	3084	1	LM1A_MOUSE
27	353	9.2	3075	1	LM1A_HUMAN
28	351	9.2	3097	1	CADN_DROME
29	347	9.1	3106	1	LM2A_MOUSE
30	333	8.7	1504	1	SLIT_DROME
31	328.5	8.6	2471	1	NTC2_HUMAN
32	328	8.6	4705	1	FAT2_DROME
33	326	8.5	3672	1	LM2L_CAEEL

34	325	8.5	3014	1	CLR1_HUMAN.
35	322.5	8.4	4590	1	FATH_HUMAN
36	319	8.3	2471	1	NTC2_RAT
37	316	8.3	2321	1	NTC3_HUMAN
38	314.5	8.2	2470	1	NTC2_MOUSE
39	312	8.2	2003	1	NTC4_HUMAN
40	311.5	8.1	2319	1	NTC3_RAT
41	311	8.1	2318	1	NTC3_MOUSE
42	311	8.1	3034	1	CLR1_MOUSE
43	310.5	8.1	1308	1	CTA4_HUMAN
44	303.5	7.9	1331	1	CTA2_HUMAN
45	302.5	7.9	1219	1	JAG1_RAT

Q9NY66 homo sapien
Q14517 homo sapien
Q9GW30 rattus norv
Q9UM47 homo sapien
Q35516 mus musculu
Q99466 homo sapien
Q91722 rattus norv
Q61982 mus musculu
Q35161 mus musculu
Q9C0A0 homo sapien
Q9UHC6 homo sapien
Q63722 rattus norv

RESULT 1
PGBM_HUMAN
ID PGBM_HUMAN STANDARD; PRT; 4391 AA.
AC P98160; Q16287; Q9H3V5;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (Perlecan) (PLC).
GN HSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI TaxID=9606;
RX Kallunki P., Tryggvason K.;
SEQUENCE FROM N.A. PubMed=1730768;
MEDLINE=92112994;
RA Kallunki P., Tryggvason K.;
RT "Human basement membrane heparan sulfate proteoglycan core protein: a
RT 467-kD protein containing multiple domains resembling elements of the
RT low density lipoprotein receptor, laminin, neural cell adhesion
RT molecules, and epidermal growth factor.";
RL J. Cell Biol. 116:559-571(1992).
[1]
[2]
SEQUENCE FROM N.A.
TISSUE=Colon, and Skin;
RX MEDLINE=92235084; PubMed=1569102;
RA Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT "Primary structure of the human heparan sulfate proteoglycan from
RT basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT domains homologous to the low density lipoprotein receptor, laminin,
RT neural cell adhesion molecules, and epidermal growth factor.";
RL J. Biol. Chem. 267:8544-8557(1992).
[3]
SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SJS1 TYR-1532.
RX MEDLINE=20553141; PubMed=1101850;
RA Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barral D.,
RA Beighton P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,
RA Samson D., Urtizberea J.A., Lehmann-Horn F., Weissenbach J.,
RA Hentati P., Fontaine B.;
RT "Perlecan, the major proteoglycan of basement membranes, is altered in
RT patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia).";
RL Nat. Genet. 26:480-483(2000).
[4]
SEQUENCE OF 1016-1470 FROM N.A.
TISSUE=Colon;
RX MEDLINE=91365376; PubMed=1679749;
RA Dodge G.R., Kovacszy I., Chu M.L., Haesell J.R., McBride O.W.,
RA Vi H.F., Iozzo R.V.;
RT "Heparan sulfate proteoglycan of human colon: partial molecular
RT cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT short arm of human chromosome 1.";
RL Genomics 10:673-680(1991).
[5]
SEQUENCE OF 890-1396 FROM N.A.
TISSUE=Fibrosarcoma;

ALIGNMENTS

RX MEDLINE=92120660; PubMed=1685141;
 RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
 RA Tryggvason K.,
 RT "cloning of human heparan sulfate proteoglycan core protein,
 RT assignment of the gene (HSPG2) to 1p36.1-p35 and identification of
 RT a BamHI restriction fragment length polymorphism.";
 RL Genomics 11:389-396 (1991).
 RN [6]
 RN SEQUENCE OF 1-21 FROM N.A.
 RP MEDLINE=94052171; PubMed=8234307;
 RX Cohen I.R., Graessels S., Murdoch A.D., Iozzo R.V.,
 RA "Structural characterization of the complete human perlecan gene and
 RT its promoter.";
 RT Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408 (1993).
 RN [7]
 RN CARBOHYDRATE-LINKAGE SITE ASN-2121.
 RP MEDLINE=22660472; PubMed=12754519;
 RX Zhang H., Li X.-J., Martin D.B., Abersold R.,
 RA "Identification and quantification of N-linked glycoproteins using
 RT hydrazide chemistry, stable isotope labeling and mass spectrometry.";
 RL Nat. Biotechnol. 21:660-666 (2003).
 CC -!- FUNCTION: This protein is an integral component of basement
 CC membranes. It is responsible for the fixed negative electrostatic
 CC charge and is involved in the charge-selective ultrafiltration
 CC properties. It serves as an attachment substrate for cells.
 CC -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
 CC dimers or stellate structures. It interacts with other basement
 CC membrane components such as laminin, prolargin and collagen type
 CC IV.
 CC -!- CELLULAR LOCATION: Extracellular.
 CC -!- TISSUE SPECIFICITY: Found in the basement membranes.
 CC -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
 CC AND O-LINKED OLIGOSACCHARIDES.
 CC -!- DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel
 CC syndrome [SJS] [MIM:25800]; a rare autosomal recessive disorder
 CC characterized by permanent myotonia (prolonged failure of muscle
 CC relaxation) and skeletal dysplasia, resulting in reduced stature,
 CC kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.
 CC -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
 CC -!- SIMILARITY: Contains 11 laminin EGF-like domains.
 CC -!- SIMILARITY: Contains 3 laminin IV domains.
 CC -!- SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.
 CC -!- SIMILARITY: Contains 3 laminin G-like domains.
 CC -!- SIMILARITY: Contains 4 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 SEA domain.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X62515; CAA44373.1; -.
 CC EMBL; M85289; AAA52700.1; -.
 CC EMBL; AL445795; CAC18534.1; -.
 CC EMBL; M64283; AAA52699.1; -.
 CC EMBL; S76436; AAB21121.2; -.
 CC EMBL; L22078; -; NOT_ANNOTATED_CDS.
 CC PIR; A38096; A38096.
 CC RSP; P00740; IEDM.
 CC Stena-2DPAGE; P98160; -.
 CC Genew; HGNC:5273; HSPG2.
 CC MIM; 142461; -.
 CC MIM; 25800; -.
 CC InterPro; IPR008985; ConA_like_lec_g1.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR006210; IEGF.
 CC InterPro; IPR007110; IG-like.
 CC InterPro; IPR003599; IG.
 CC InterPro; IPR003598; Ig_c2.

DR InterPro; IPR003596; Ig_v.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR000082; SEA_domain.
 DR Pfam; PF00008; EGF; 4.
 DR Pfam; PF00047; IG; 22.
 DR Pfam; PF00052; laminin_B; 3.
 DR Pfam; PF00053; laminin_EGF; 7.
 DR Pfam; PF00054; laminin_G; 3.
 DR Pfam; PF00057; ldl_recept_a; 4.
 DR Pfam; PF01390; SEA; 1.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR PRODOM; PD003031; Laminin_B; 3.
 DR SMART; SM00181; EGF; 15.
 DR SMART; SM00180; EGF_Lam; 12.
 DR SMART; SM00409; IG; 22.
 DR SMART; SM00408; IGc2; 21.
 DR SMART; SM00406; IGv; 7.
 DR SMART; SM00281; LamB; 3.
 DR SMART; SM00282; LamG; 3.
 DR SMART; SM00192; LDLa; 4.
 DR SMART; SM00200; SEA; 1.
 DR PROSITE; PS00022; EGF_1; 9.
 DR PROSITE; PS01186; EGF_2; 6.
 DR PROSITE; PS00026; EGF_3; 4.
 DR PROSITE; PS00835; IG_LIKE; 22.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
 DR PROSITE; PS01209; LDLRA_1; 4.
 DR PROSITE; PS00068; LDLRA_2; 4.
 DR PROSITE; PS00024; SEA; 1.
 DR Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
 KW Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
 KW Extracellular matrix; EGF-like domain; Disease mutation.
 FT SIGNAL 1 21
 FT CHAIN 22 4391
 FT DOMAIN 80 194
 FT DOMAIN 198 235
 FT DOMAIN 284 320
 FT DOMAIN 324 360
 FT DOMAIN 367 404
 FT DOMAIN 405 504
 FT DOMAIN 521 530
 FT DOMAIN 531 730
 FT DOMAIN 731 763
 FT DOMAIN 764 813
 FT DOMAIN 814 871
 FT DOMAIN 879 933
 FT DOMAIN 924 933
 FT DOMAIN 934 1125
 FT DOMAIN 1126 1158
 FT DOMAIN 1159 1208
 FT DOMAIN 1209 1265
 FT DOMAIN 1275 1324
 FT DOMAIN 1325 1334
 FT DOMAIN 1335 1529
 FT DOMAIN 1530 1562
 FT DOMAIN 1563 1612
 FT DOMAIN 1613 1670
 FT DOMAIN 1672 1771
 FT DOMAIN 1771 1865
 FT DOMAIN 1866 1955
 FT DOMAIN 1956 2051
 FT DOMAIN 2052 2151
 FT DOMAIN 2152 2244
 FT DOMAIN 2245 2340
 FT DOMAIN 2341 2436
 FT DOMAIN 2437 2533
 FT DOMAIN 2534 2629
 FT DOMAIN 2630 2726
 FT DOMAIN 2726 2826
 FT DOMAIN 2827 2926
 FT DOMAIN 2927 3026
 FT DOMAIN 3027 3126
 FT DOMAIN 3127 3226
 FT DOMAIN 3227 3326
 FT DOMAIN 3327 3426
 FT DOMAIN 3427 3526
 FT DOMAIN 3527 3626
 FT DOMAIN 3627 3726
 FT DOMAIN 3727 3826
 FT DOMAIN 3827 3926
 FT DOMAIN 3927 4026
 FT DOMAIN 4027 4126
 FT DOMAIN 4127 4226
 FT DOMAIN 4227 4326
 FT DOMAIN 4327 4426
 FT DOMAIN 4427 4526
 FT DOMAIN 4527 4626
 FT DOMAIN 4627 4726
 FT DOMAIN 4727 4826
 FT DOMAIN 4827 4926
 FT DOMAIN 4927 5026
 FT DOMAIN 5027 5126
 FT DOMAIN 5127 5226
 FT DOMAIN 5227 5326
 FT DOMAIN 5327 5426
 FT DOMAIN 5427 5526
 FT DOMAIN 5527 5626
 FT DOMAIN 5627 5726
 FT DOMAIN 5727 5826
 FT DOMAIN 5827 5926
 FT DOMAIN 5927 6026
 FT DOMAIN 6027 6126
 FT DOMAIN 6127 6226
 FT DOMAIN 6227 6326
 FT DOMAIN 6327 6426
 FT DOMAIN 6427 6526
 FT DOMAIN 6527 6626
 FT DOMAIN 6627 6726
 FT DOMAIN 6727 6826
 FT DOMAIN 6827 6926
 FT DOMAIN 6927 7026
 FT DOMAIN 7027 7126
 FT DOMAIN 7127 7226
 FT DOMAIN 7227 7326
 FT DOMAIN 7327 7426
 FT DOMAIN 7427 7526
 FT DOMAIN 7527 7626
 FT DOMAIN 7627 7726
 FT DOMAIN 7727 7826
 FT DOMAIN 7827 7926
 FT DOMAIN 7927 8026
 FT DOMAIN 8027 8126
 FT DOMAIN 8127 8226
 FT DOMAIN 8227 8326
 FT DOMAIN 8327 8426
 FT DOMAIN 8427 8526
 FT DOMAIN 8527 8626
 FT DOMAIN 8627 8726
 FT DOMAIN 8727 8826
 FT DOMAIN 8827 8926
 FT DOMAIN 8927 9026
 FT DOMAIN 9027 9126
 FT DOMAIN 9127 9226
 FT DOMAIN 9227 9326
 FT DOMAIN 9327 9426
 FT DOMAIN 9427 9526
 FT DOMAIN 9527 9626
 FT DOMAIN 9627 9726
 FT DOMAIN 9727 9826
 FT DOMAIN 9827 9926
 FT DOMAIN 9927 10026
 FT DOMAIN 10027 10126
 FT DOMAIN 10127 10226
 FT DOMAIN 10227 10326
 FT DOMAIN 10327 10426
 FT DOMAIN 10427 10526
 FT DOMAIN 10527 10626
 FT DOMAIN 10627 10726
 FT DOMAIN 10727 10826
 FT DOMAIN 10827 10926
 FT DOMAIN 10927 11026
 FT DOMAIN 11027 11126
 FT DOMAIN 11127 11226
 FT DOMAIN 11227 11326
 FT DOMAIN 11327 11426
 FT DOMAIN 11427 11526
 FT DOMAIN 11527 11626
 FT DOMAIN 11627 11726
 FT DOMAIN 11727 11826
 FT DOMAIN 11827 11926
 FT DOMAIN 11927 12026
 FT DOMAIN 12027 12126
 FT DOMAIN 12127 12226
 FT DOMAIN 12227 12326
 FT DOMAIN 12327 12426
 FT DOMAIN 12427 12526
 FT DOMAIN 12527 12626
 FT DOMAIN 12627 12726
 FT DOMAIN 12727 12826
 FT DOMAIN 12827 12926
 FT DOMAIN 12927 13026
 FT DOMAIN 13027 13126
 FT DOMAIN 13127 13226
 FT DOMAIN 13227 13326
 FT DOMAIN 13327 13426
 FT DOMAIN 13427 13526
 FT DOMAIN 13527 13626
 FT DOMAIN 13627 13726
 FT DOMAIN 13727 13826
 FT DOMAIN 13827 13926
 FT DOMAIN 13927 14026
 FT DOMAIN 14027 14126
 FT DOMAIN 14127 14226
 FT DOMAIN 14227 14326
 FT DOMAIN 14327 14426
 FT DOMAIN 14427 14526
 FT DOMAIN 14527 14626
 FT DOMAIN 14627 14726
 FT DOMAIN 14727 14826
 FT DOMAIN 14827 14926
 FT DOMAIN 14927 15026
 FT DOMAIN 15027 15126
 FT DOMAIN 15127 15226
 FT DOMAIN 15227 15326
 FT DOMAIN 15327 15426
 FT DOMAIN 15427 15526
 FT DOMAIN 15527 15626
 FT DOMAIN 15627 15726
 FT DOMAIN 15727 15826
 FT DOMAIN 15827 15926
 FT DOMAIN 15927 16026
 FT DOMAIN 16027 16126
 FT DOMAIN 16127 16226
 FT DOMAIN 16227 16326
 FT DOMAIN 16327 16426
 FT DOMAIN 16427 16526
 FT DOMAIN 16527 16626
 FT DOMAIN 16627 16726
 FT DOMAIN 16727 16826
 FT DOMAIN 16827 16926
 FT DOMAIN 16927 17026
 FT DOMAIN 17027 17126
 FT DOMAIN 17127 17226
 FT DOMAIN 17227 17326
 FT DOMAIN 17327 17426
 FT DOMAIN 17427 17526
 FT DOMAIN 17527 17626
 FT DOMAIN 17627 17726
 FT DOMAIN 17727 17826
 FT DOMAIN 17827 17926
 FT DOMAIN 17927 18026
 FT DOMAIN 18027 18126
 FT DOMAIN 18127 18226
 FT DOMAIN 18227 18326
 FT DOMAIN 18327 18426
 FT DOMAIN 18427 18526
 FT DOMAIN 18527 18626
 FT DOMAIN 18627 18726
 FT DOMAIN 18727 18826
 FT DOMAIN 18827 18926
 FT DOMAIN 18927 19026
 FT DOMAIN 19027 19126
 FT DOMAIN 19127 19226
 FT DOMAIN 19227 19326
 FT DOMAIN 19327 19426
 FT DOMAIN 19427 19526
 FT DOMAIN 19527 19626
 FT DOMAIN 19627 19726
 FT DOMAIN 19727 19826
 FT DOMAIN 19827 19926
 FT DOMAIN 19927 20026
 FT DOMAIN 20027 20126
 FT DOMAIN 20127 20226
 FT DOMAIN 20227 20326
 FT DOMAIN 20327 20426
 FT DOMAIN 20427 20526
 FT DOMAIN 20527 20626
 FT DOMAIN 20627 20726
 FT DOMAIN 20727 20826
 FT DOMAIN 20827 20926
 FT DOMAIN 20927 21026
 FT DOMAIN 21027 21126
 FT DOMAIN 21127 21226
 FT DOMAIN 21227 21326
 FT DOMAIN 21327 21426
 FT DOMAIN 21427 21526
 FT DOMAIN 21527 21626
 FT DOMAIN 21627 21726
 FT DOMAIN 21727 21826
 FT DOMAIN 21827 21926
 FT DOMAIN 21927 22026
 FT DOMAIN 22027 22126
 FT DOMAIN 22127 22226
 FT DOMAIN 22227 22326
 FT DOMAIN 22327 22426
 FT DOMAIN 22427 22526
 FT DOMAIN 22527 22626
 FT DOMAIN 22627 22726
 FT DOMAIN 22727 22826
 FT DOMAIN 22827 22926
 FT DOMAIN 22927 23026
 FT DOMAIN 23027 23126
 FT DOMAIN 23127 23226
 FT DOMAIN 23227 23326
 FT DOMAIN 23327 23426
 FT DOMAIN 23427 23526
 FT DOMAIN 23527 23626
 FT DOMAIN 23627 23726
 FT DOMAIN 23727 23826
 FT DOMAIN 23827 23926
 FT DOMAIN 23927 24026
 FT DOMAIN 24027 24126
 FT DOMAIN 24127 24226
 FT DOMAIN 24227 24326
 FT DOMAIN 24327 24426
 FT DOMAIN 24427 24526
 FT DOMAIN 24527 24626
 FT DOMAIN 24627 24726
 FT DOMAIN 24727 24826
 FT DOMAIN 24827 24926
 FT DOMAIN 24927 25026
 FT DOMAIN 25027 25126
 FT DOMAIN 25127 25226
 FT DOMAIN 25227 25326
 FT DOMAIN 25327 25426
 FT DOMAIN 25427 25526
 FT DOMAIN 25527 25626
 FT DOMAIN 25627 25726
 FT DOMAIN 25727 25826
 FT DOMAIN 25827 25926
 FT DOMAIN 25927 26026
 FT DOMAIN 26027 26126
 FT DOMAIN 26127 26226
 FT DOMAIN 26227 26326
 FT DOMAIN 26327 26426
 FT DOMAIN 26427 26526
 FT DOMAIN 26527 26626
 FT DOMAIN 26627 26726
 FT DOMAIN 26727 26826
 FT DOMAIN 26827 26926
 FT DOMAIN 26927 27026
 FT DOMAIN 27027 27126
 FT DOMAIN 27127 27226
 FT DOMAIN 27227 27326
 FT DOMAIN 27327 27426
 FT DOMAIN 27427 27526
 FT DOMAIN 27527 27626
 FT DOMAIN 27627 27726
 FT DOMAIN 27727 27826
 FT DOMAIN 27827 27926
 FT DOMAIN 27927 28026
 FT DOMAIN 28027 28126
 FT DOMAIN 28127 28226
 FT DOMAIN 28227 28326
 FT DOMAIN 28327 28426
 FT DOMAIN 28427 28526
 FT DOMAIN 28527 28626
 FT DOMAIN 28627 28726
 FT DOMAIN 28727 28826
 FT DOMAIN 28827 28926
 FT DOMAIN 28927 29026
 FT DOMAIN 29027 29126
 FT DOMAIN 29127 29226
 FT DOMAIN 29227 29326
 FT DOMAIN 29327 29426
 FT DOMAIN 29427 29526
 FT DOMAIN 29527 29626
 FT DOMAIN 29627 29726
 FT DOMAIN 29727 29826
 FT DOMAIN 29827 29926
 FT DOMAIN 29927 30026
 FT DOMAIN 30027 30126
 FT DOMAIN 30127 30226
 FT DOMAIN 30227 30326
 FT DOMAIN 30327 30426
 FT DOMAIN 30427 30526
 FT DOMAIN 30527 30626
 FT DOMAIN 30627 30726
 FT DOMAIN 30727 30826
 FT DOMAIN 30827 30926
 FT DOMAIN 30927 31026
 FT DOMAIN 31027 31126
 FT DOMAIN 31127 31226
 FT DOMAIN 31227 31326
 FT DOMAIN 31327 31426
 FT DOMAIN 31427 31526
 FT DOMAIN 31527 31626
 FT DOMAIN 31627 31726
 FT DOMAIN 31727 31826
 FT DOMAIN 31827 31926
 FT DOMAIN 31927 32026
 FT DOMAIN 32027 32126
 FT DOMAIN 32127 32226
 FT DOMAIN 32227 32326
 FT DOMAIN 32327 32426
 FT DOMAIN 32427 32526
 FT DOMAIN 32527 32626
 FT DOMAIN 32627 32726
 FT DOMAIN 32727 32826
 FT DOMAIN 32827 32926
 FT DOMAIN 32927 33026
 FT DOMAIN 33027 33126
 FT DOMAIN 33127 33226
 FT DOMAIN 33227 33326
 FT DOMAIN 33327 33426
 FT DOMAIN 33427 33526
 FT DOMAIN 33527 33626
 FT DOMAIN 33627 33726
 FT DOMAIN 33727 33826
 FT DOMAIN 33827 33926
 FT DOMAIN 33927 34026
 FT DOMAIN 34027 34126
 FT DOMAIN 34127 34226
 FT DOMAIN 34227 34326
 FT DOMAIN 34327 34426
 FT DOMAIN 34427 34526
 FT DOMAIN 34527 34626
 FT DOMAIN 34627 34726
 FT DOMAIN 34727 34826
 FT DOMAIN 34827 34926
 FT DOMAIN 34927 35026
 FT DOMAIN 35027 35126
 FT DOMAIN 35127 35226
 FT DOMAIN 35227 35326
 FT DOMAIN 35327 35426
 FT DOMAIN 35427 35526
 FT DOMAIN 35527 35626
 FT DOMAIN 35627 35726
 FT DOMAIN 35727 35826
 FT DOMAIN 35827 35926
 FT DOMAIN 35927 36026
 FT DOMAIN 36027 36126
 FT DOMAIN 36127 36226
 FT DOMAIN 36227 36326
 FT DOMAIN 36327 36426
 FT DOMAIN 36427 36526
 FT DOMAIN 36527 36626
 FT DOMAIN 36627 36726
 FT DOMAIN 36727 36826
 FT DOMAIN 36827 36926
 FT DOMAIN 36927 37026
 FT DOMAIN 37027 37126
 FT DOMAIN 37127 37226
 FT DOMAIN 37227 37326
 FT DOMAIN 37327 37426
 FT DOMAIN 37427 37526
 FT DOMAIN 37527 37626
 FT DOMAIN 37627 37726
 FT DOMAIN 37727 37826
 FT DOMAIN 37827 37926
 FT DOMAIN 37927 38026
 FT DOMAIN 38027 38126
 FT DOMAIN 38127 38226
 FT DOMAIN 38227 38326
 FT DOMAIN 38327 38426
 FT DOMAIN 38427 38526
 FT DOMAIN 38527 38626
 FT DOMAIN 38627 38726
 FT DOMAIN 38727 38826
 FT DOMAIN 38827 38926
 FT DOMAIN 38927 39026
 FT DOMAIN 39027 39126
 FT DOMAIN 39127 39226
 FT DOMAIN 39227 39326
 FT DOMAIN 39327 39426
 FT DOMAIN 39427 39526
 FT DOMAIN 39527 39626
 FT DOMAIN 39627 39726
 FT DOMAIN 39727 39826
 FT DOMAIN 39827 39926
 FT DOMAIN 39927 40026
 FT DOMAIN 40027 40126
 FT DOMAIN 40127 40226
 FT DOMAIN 40227 40326
 FT DOMAIN 40327 40426
 FT DOMAIN 40427 40526
 FT DOMAIN 40527 40626
 FT DOMAIN 40627 40726
 FT DOMAIN 40727 40826
 FT DOMAIN 40827 40926
 FT DOMAIN 40927 41026
 FT DOMAIN 41027 41126
 FT DOMAIN 41127 41226
 FT DOMAIN 41227 41326
 FT DOMAIN 41327 41426
 FT DOMAIN 41427 41526
 FT DOMAIN 41527 41626
 FT DOMAIN 41627 41726
 FT DOMAIN 41727 41826
 FT DOMAIN 41827 41926
 FT DOMAIN 41927 42026
 FT DOMAIN 42027 42126
 FT DOMAIN 42127 42226
 FT DOMAIN 42227 42326
 FT DOMAIN 42327 42426
 FT DOMAIN 42427 42526
 FT DOMAIN 42527 42626
 FT DOMAIN 42627 42726
 FT DOMAIN 42727 42826
 FT DOMAIN 42827 42926
 FT DOMAIN 42927 43026
 FT DOMAIN 43027 43126
 FT DOMAIN 43127 43226
 FT DOMAIN 43227 43326
 FT DOMAIN 43327 43426
 FT DOMAIN 43427 43526
 FT DOMAIN 43527 43626
 FT DOMAIN 43627 43726
 FT DOMAIN 43727 43826
 FT DOMAIN 43827 43926
 FT DOMAIN 43927 44026
 FT DOMAIN 44027 44126
 FT DOMAIN 44127 44226
 FT DOMAIN 44227 44326
 FT DOMAIN 44327 44426
 FT DOMAIN 44427 44526
 FT DOMAIN 44527 44626
 FT DOMAIN 44627 44726
 FT DOMAIN 44727 44826
 FT DOMAIN 44827 44926
 FT DOMAIN 44927 45026
 FT DOMAIN 45027 45126
 FT DOMAIN 45127 45226
 FT DOMAIN 45227 45326
 FT DOMAIN 45327 45426
 FT DOMAIN 45427 45526
 FT DOMAIN 45527 45626
 FT DOMAIN 45627 45726
 FT DOMAIN 45727 45826
 FT DOMAIN 45827 45926
 FT DOMAIN 45927 46026
 FT DOMAIN 46027 46126
 FT DOMAIN 46127 46226
 FT DOMAIN 46227 46326
 FT DOMAIN 46327 46426
 FT DOMAIN 46427 46526
 FT DOMAIN 46527 46626
 FT DOMAIN 46627 46726
 FT DOMAIN 46727 46826
 FT DOMAIN 46827 46926
 FT DOMAIN 46927 47026
 FT DOMAIN 47027 47126
 FT DOMAIN 47127 47226
 FT DOMAIN 47227 47326
 FT DOMAIN 47327 47426
 FT DOMAIN 47427 47526
 FT DOMAIN 47527 47626
 FT DOMAIN 47627 47726
 FT DOMAIN 47727 47826
 FT DOMAIN 47827 47926
 FT DOMAIN 47927 48026
 FT DOMAIN 48027 48126
 FT DOMAIN 48127 48226
 FT DOMAIN 48227 48326
 FT DOMAIN 48327 48426
 FT DOMAIN 48427 48526
 FT DOMAIN 48527 48626
 FT DOMAIN 48627 48726
 FT DOMAIN 48727 48826
 FT DOMAIN 48827 48926
 FT DOMAIN 48927 49026
 FT DOMAIN 49027 49126
 FT DOMAIN 49127 49226
 FT DOMAIN 49227 49326
 FT DOMAIN 49327 49426
 FT DOMAIN 49427 49526
 FT DOMAIN 49527 49626
 FT DOMAIN 49627 49726
 FT DOMAIN 49727 49826
 FT DOMAIN 49827 49926
 FT DOMAIN 49927 50026
 FT DOMAIN 50027 50126
 FT DOMAIN 50127 50226
 FT DOMAIN 50227 50326
 FT DOMAIN 50327 50426
 FT DOMAIN 50427 50526
 FT DOMAIN 50527 50626
 FT DOMAIN 50627 50726
 FT DOMAIN 50727 50826
 FT DOMAIN 50827 50926
 FT DOMAIN 50927 51026
 FT DOMAIN 51027 51126
 FT DOMAIN 51127 51226
 FT DOMAIN 51227 51326
 FT DOMAIN 51327 51426
 FT DOMAIN 51427 51526
 FT DOMAIN 51527 51626
 FT DOMAIN 51627 51726
 FT DOMAIN 51727 51826
 FT DOMAIN 51827 51926
 FT DOMAIN 51927 52026
 FT DOMAIN 52027 52126
 FT DOMAIN 52127 52226
 FT DOMAIN 52227 52326
 FT DOMAIN 52327 52426
 FT DOMAIN 52427 52526
 FT DOMAIN 52527 52626
 FT DOMAIN 52627 52726
 FT DOMAIN 52727 52826
 FT DOMAIN 52827 52926
 FT DOMAIN 52927 53026
 FT DOMAIN 53027 53126
 FT DOMAIN 53127 53226
 FT DOMAIN 53227 53326
 FT DOMAIN 53327 53426
 FT DOMAIN 53427 53526
 FT DOMAIN 53527 53626
 FT DOMAIN 53627 53726
 FT DOMAIN 53727 53826
 FT DOMAIN 53827 53926
 FT DOMAIN 53927 54026
 FT DOMAIN 54027 54126
 FT DOMAIN 54127 54226
 FT DOMAIN 54227 54326
 FT DOMAIN 54327 54426
 FT DOMAIN 54427 54526
 FT DOMAIN 54527 54626
 FT DOMAIN 54627 54726
 FT DOMAIN 54727 54826
 FT DOMAIN 54827 54926
 FT DOMAIN 54927 55026
 FT DOMAIN 55027 55126
 FT DOMAIN 55127 55226
 FT DOMAIN 55227 55326
 FT DOMAIN 553

Query Match 100.0%; Score 3825; DB 1; Length 4391;
Best Local Similarity 100.0%; Pred. NO. 1.3e-233;
Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EIKITRPSADGMLYNGKVPQSPTNLNRPDFTSFLVGGRRPEFRDAGSGWATI 60
DB 3687 EIKITRPSADGMLYNGKVPQSPTNLNRPDFTSFLVGGRRPEFRDAGSGWATI 3746
QY 61 RHTPTALGHFHTVTLRLSLTQSLIVGLDLPVNGTSQKFGDLNLELYGYPIYGA 120
DB 3747 RHTPTALGHFHTVTLRLSLTQSLIVGLDLPVNGTSQKFGDLNLELYGYPIYGA 3806
QY 121 IPKAGLSGGFICVRLRIGQBEIVPHDLNLTAKHSHCPICRDPCCNGQCHDSRSS 180
DB 3807 IPKAGLSGGFICVRLRIGQBEIVPHDLNLTAKHSHCPICRDPCCNGQCHDSRSS 3866
QY 181 YVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRHLGRSLRCESGVT 240
DB 3867 YVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRHLGRSLRCESGVT 3926
QY 241 VTPPISGAGSYLALPALNTWHLRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
DB 3927 VTPPISGAGSYLALPALNTWHLRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 3986
QY 301 GHLEFRELGLAVLRSAPLALGHWHRVSAERLNKDGSLRVNGRPVLRSSPKSQGL 360
DB 3987 GHLEFRELGLAVLRSAPLALGHWHRVSAERLNKDGSLRVNGRPVLRSSPKSQGL 4046
QY 361 NLHTLLYLGVEVPVLSPATNWSAHFRGCVGVSVNGKRLDLTYSFLGSGIQCCYDSS 420
DB 4047 NLHTLLYLGVEVPVLSPATNWSAHFRGCVGVSVNGKRLDLTYSFLGSGIQCCYDSS 4106
QY 421 PCERQPCQHGATCPMAGEYEFQCLRDGPKGDLCSHEENPCOLRPPCLHGTCGTRCLC 480
DB 4107 PCERQPCQHGATCPMAGEYEFQCLRDGPKGDLCSHEENPCOLRPPCLHGTCGTRCLC 4166
QY 481 LPFGSPRCQSGHGIASDWHLESGGNDAPGQVGFYHDDGLAPFPFHVSRLSPEV 540
DB 4167 LPFGSPRCQSGHGIASDWHLESGGNDAPGQVGFYHDDGLAPFPFHVSRLSPEV 4226
QY 541 PETIELEVRTSTASGLLLQGVVEVAGQKDFISLGLQDGHVLPYQLSGGEARLVSED 600
DB 4227 PETIELEVRTSTASGLLLQGVVEVAGQKDFISLGLQDGHVLPYQLSGGEARLVSED 4286
QY 601 PINDGEWHRTALREGRGSIQVDGEELVSGRSPGNVAVNAKGSVYIGGAPDVATLTGG 660
DB 4287 PINDGEWHRTALREGRGSIQVDGEELVSGRSPGNVAVNAKGSVYIGGAPDVATLTGG 4346
QY 661 RFSSGITGCVKNLVLSARPGAPPPPOPLDQHQAGANTRPCPS 705
DB 4347 RFSSGITGCVKNLVLSARPGAPPPPOPLDQHQAGANTRPCPS 4391

RESULT 2
ID PGEM MOUSE STANDARD; PRT: 3707 AA.
AC Q0573;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (Perlecan) (PLC).
GN HSPG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=92078153; PubMed=1744087;
RA Noonan D.M., Fuller A., Valente P., Cai S., Horgan E., Sasaki M.,

Yamada Y., Hassell J.R.;
"The complete sequence of perlecan, a basement membrane heparan sulfate proteoglycan, reveals extensive similarity with laminin A chain, low density lipoprotein-receptor, and the neural cell adhesion molecule.";
J. Biol. Chem. 266:22939-22947(1991).
[2]
SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
MEDLINE=89034110; PubMed=2972708;
RX Noonan D.M., Horgan E.A., Ledbetter S.R., Vogeli G., Sasaki M., Yamada Y., Hassell J.R.;
"Identification of cDNA clones encoding different domains of the basement membrane heparan sulfate proteoglycan.";
J. Biol. Chem. 263:16379-16387(1988).
CC -!- FUNCTION: This protein is an integral component of basement membranes. It is responsible for the fixed negative electrostatic charge and is involved in the charge-selective ultrafiltration properties. It serves as an attachment substrate for cells.
CC -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in dimers or stellate structures. It interacts with other basement membrane components such as laminin, prolargin and collagen type IV.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Found in the basement membranes.
CC -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC -!- SIMILARITY: Contains 11 laminin EGF-like domains.
CC -!- SIMILARITY: Contains 3 laminin IV domains.
CC -!- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 3 laminin G-like domains.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 1 SEA domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: W77174; AAA39911.1; -;
EMBL: J04054; AAA39899.1; -;
EMBL: J04055; AAA39912.1; -;
PIR: S18252; S18252.
PDB: 1GL4; 28-NOV-01.
DR MGD: MGI:96257; Hspg2.
DR GO: 0005604; C:basement membrane; IDA.
DR GO: 0008104; P:protein localization; IMP.
DR InterPro: IPR008985; C:alpha-like_lec_gl.
DR InterPro: IPR000742; EGF 2.
DR InterPro: IPR006209; EGF like.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003598; Ig C2.
DR InterPro: IPR000034; Laminin B.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR002172; LDL_receptor_A.
DR Pfam: PF00008; EGF; 4.
DR Pfam: PF00047; Ig; 15.
DR Pfam: PF00052; laminin_B; 3.
DR Pfam: PF00053; laminin_EGF; 7.
DR Pfam: PF00054; laminin_G; 3.
DR Pfam: PF00057; ldl_recept_a; 4.
DR Pfam: PF01390; SEA; 1.
DR PRINTS: PR00281; LDLRECEPTOR.
DR ProDom: PD003031; Laminin_B; 3.
DR SMART: SM00180; EGF_Lam; 7.
DR SMART: SM00408; IGC2; 14.
DR SMART: SM00281; LamB; 3.
DR SMART: SM00282; LamG; 3.

DR SMART; SM00192; LDLA; 4.
 DR SMART; SM00200; SEA; 1.
 DR PROSITE; PS00022; EGF 1; 8.
 DR PROSITE; PS01186; EGF 2; 5.
 DR PROSITE; PS00026; EGF 3; 4.
 DR PROSITE; PS00835; IG-LIKE; 15.
 DR PROSITE; PS00025; LAMININ G-DOMAIN; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
 DR PROSITE; PS01209; LDLRA 1; 4.
 DR PROSITE; PS00068; LDLRA 2; 4.
 DR PROSITE; PS00024; SEA; 1.
 KW Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
 KW Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
 KW Extracellular matrix; EGF-like domain; 3D-structure.
 FT SIGNAL 1 21
 FT CHAIN 22 3707
 FT DOMAIN 80 194
 FT DOMAIN 195 234
 FT DOMAIN 281 319
 FT DOMAIN 320 359
 FT DOMAIN 360 403
 FT DOMAIN 404 504
 FT DOMAIN 521 530
 FT DOMAIN 531 730
 FT DOMAIN 731 763
 FT DOMAIN 764 813
 FT DOMAIN 814 871
 FT DOMAIN 879 923
 FT DOMAIN 924 933
 FT DOMAIN 934 1125
 FT DOMAIN 1126 1158
 FT DOMAIN 1159 1208
 FT DOMAIN 1209 1265
 FT DOMAIN 1275 1324
 FT DOMAIN 1325 1334
 FT DOMAIN 1335 1529
 FT DOMAIN 1530 1562
 FT DOMAIN 1563 1612
 FT DOMAIN 1613 1670
 FT DOMAIN 1671 1771
 FT DOMAIN 1772 1865
 FT DOMAIN 1866 1954
 FT DOMAIN 1955 2049
 FT DOMAIN 2050 2148
 FT DOMAIN 2149 2244
 FT DOMAIN 2245 2343
 FT DOMAIN 2344 2436
 FT DOMAIN 2437 2532
 FT DOMAIN 2533 2619
 FT DOMAIN 2620 2720
 FT DOMAIN 2721 2809
 FT DOMAIN 2810 2895
 FT DOMAIN 2896 2980
 FT DOMAIN 2984 3162
 FT DOMAIN 3163 3241
 FT DOMAIN 3245 3425
 FT DOMAIN 3518 3705
 FT SITE 65 67
 FT SITE 71 73
 FT SITE 76 78
 FT SITE 3617 3617
 FT DISULFID 199 212
 FT DISULFID 206 225
 FT DISULFID 219 234
 FT DISULFID 285 297
 FT DISULFID 292 310
 FT DISULFID 304 319
 FT DISULFID 325 337
 FT DISULFID 332 350
 FT DISULFID 344 359
 FT DISULFID 368 381

FT DISULFID 375 394 BY SIMILARITY.
 FT DISULFID 388 403 BY SIMILARITY.
 FT DISULFID 428 479 BY SIMILARITY.
 FT DISULFID 764 773 BY SIMILARITY.
 FT DISULFID 766 780 BY SIMILARITY.
 FT DISULFID 783 792 BY SIMILARITY.
 FT DISULFID 795 811 BY SIMILARITY.
 FT DISULFID 814 829 BY SIMILARITY.
 FT DISULFID 816 839 BY SIMILARITY.
 FT DISULFID 842 851 BY SIMILARITY.
 FT DISULFID 854 869 BY SIMILARITY.
 FT DISULFID 1159 1168 BY SIMILARITY.
 FT DISULFID 1161 1175 BY SIMILARITY.
 FT DISULFID 1178 1187 BY SIMILARITY.
 FT DISULFID 1190 1206 BY SIMILARITY.
 FT DISULFID 1209 1224 BY SIMILARITY.
 FT DISULFID 1211 1234 BY SIMILARITY.
 FT DISULFID 1237 1246 BY SIMILARITY.
 FT DISULFID 1249 1263 BY SIMILARITY.
 FT DISULFID 1275 1287 BY SIMILARITY.
 FT DISULFID 1277 1293 BY SIMILARITY.
 FT DISULFID 1295 1304 BY SIMILARITY.
 FT DISULFID 1307 1322 BY SIMILARITY.
 FT DISULFID 1563 1572 BY SIMILARITY.
 FT DISULFID 1565 1579 BY SIMILARITY.
 FT DISULFID 1582 1591 BY SIMILARITY.
 FT DISULFID 1594 1610 BY SIMILARITY.
 FT DISULFID 1613 1628 BY SIMILARITY.
 FT DISULFID 1641 1650 BY SIMILARITY.
 FT DISULFID 1653 1668 BY SIMILARITY.
 FT DISULFID 1792 1839 BY SIMILARITY.
 FT DISULFID 1886 1932 BY SIMILARITY.
 FT DISULFID 1976 2021 BY SIMILARITY.
 FT DISULFID 2073 2118 BY SIMILARITY.
 FT DISULFID 2170 2215 BY SIMILARITY.
 FT DISULFID 2268 2313 BY SIMILARITY.

Query Match

88.6%; Score 3389.5; DB 1; Length 3707;

Best Local Similarity 88.9%; Pred. No. 3.6e-206;

Matches 627; Conservative 33; Mismatches 40; Indels 5; Gaps 3;

QY 1 EIKITFRPDSADGMLLYNQKRVPGSPNTNLANRQPDFISFGLVGRPEFRFDAGSGMATI 60
 DB 3008 EIKITFRPDSADGMLLYNQKRVPGSPNTNLANRQPDFISFGLVGRPEFRFDAGSGMATI 3064
 QY 61 RPTPLALCHPHTVTLRLSLTQGSLLVGDLPVNGTSQKFGGLDNEELYLGYPDYGA 120
 DB 3065 RPTPLALCHPHTVTLRLSLTQGSLLVGDLPVNGTSQKFGGLDNEELYLGYPDYGA 3124
 QY 121 IPKAGLSGFTGCVRELRIQGEIVFHDNLNTHGISHCPTCRDRPCQNGGQCHDSSES 180
 DB 3125 IPKAGLSGFTGCVRELRIQGEIVFHDNLNTHGISHCPTCRDRPCQNGGQCHDSSES 3184
 QY 181 YVCVCPAGFTGSRCHESQALHCHPEAGCPDATCVNRPDGRGYTCRCHLGRSGLRCEGVT 240
 DB 3185 YVCVCPAGFTGSRCHESQALHCHPEAGCPDATCVNRPDGRGYTCRCHLGRSGLRCEGVT 3243
 QY 241 VTTPLSGAGSVLALPALTNTNTHRLDVEFKPLAPDGVLLFSGKSGPVDFVSLAWVG 300
 DB 3244 VTTPLSGAGSVLALPALTNTNTHRLDVEFKPLAPDGVLLFSGKSGPVDFVSLAWVG 3303
 QY 301 GHLEFRYELGSLAVLRSAPALGRVHRVSAERLNKDGSLRVNGRPVLRSSPGKSQGL 360
 DB 3304 GHLEFRYELGSLAVLRSAPALGRVHRVSAERLNKDGSLRVNGRPVLRSSPGKSQGL 3363
 QY 361 NLHTLLYLGVEPSPVPLSPATNMSAHFRCVGVSVNGKRLDITYPLSGSGGQGVYDSS 420
 DB 3364 NLHTLLYLGVEPSPVPLSPATNMSAHFRCVGVSVNGKRLDITYPLSGSGGQGVYDSS 3423
 QY 421 PCERQPCQHGATCMPAGEYEFQCLCRDPFGKDLCEHENPCQLREPCQLHGTCQGTCLC 480
 DB 3424 PCERQPCQHGATCMPAGEYEFQCLCRDPFGKDLCEHENPCQLREPCQLHGTCQGTCLC 3483

QY 481 LPFGSPRCQGGHGIABSDWHLEGGGNDAPGQYGFYHDDGFLAPFFHVSRLPEV 540
 Db 3484 LPFGSPRCQGGHGIABSDWHLEGGGNDAPGQYGFYHDDGFLAPFFHVSRLPEV 3543
 QY 541 PETIELEVRSTASGLLLQGVVEGAGOGKDFISLGLQDGLHVPYQLGSGEARLYSED 600
 Db 3544 PETIELEVRSTADGLLLQGVV-VRASRSKDFISLGLQDGLHVPYQLGSGEARLYSGD 3602
 QY 601 PINDGEHVRVTLREGRSISQVDSBELVSGSPGPNVAVNAKGVYTGAPDVATLTGG 660
 Db 3603 PINDGEHVRVTLREGRSISQVDSBELVSGSPGPNVAVNAKGVYTGAPDVATLTGG 3662
 QY 661 RFSSGIRGCVKNLVLSHSPGPPPOPLDQHRAGAGANTPCPS 705
 Db 3663 KFSSGIRGCVKNLVLSHSPGPPPOPLDQHRAGAGANTPCPS 3707

RESULT 3
 AGRI RAT
 ID AGRI RAT STANDARD; PRT; 1959 AA.
 AC P25304; Q63034;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Agtrin precursor.
 GN AGRN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE OF 1-1779 AND 1799-1959 FROM N.A.
 RC TISSUE=Embryonic spinal cord;
 RX MEDLINE=91222570; PubMed=181019;
 RA Rupp F., Payan D.G., Magill-Solc C., Cowan D.M., Scheller R.H.;
 RT "Structure and expression of a rat agrin.";
 RL Neuron 6:811-823(1991).
 RN [2]
 RP SEQUENCE OF 1777-1801 FROM N.A.
 RX MEDLINE=92407628; PubMed=1326608;
 RA Rupp F., Oezcelik T., Linial M., Peterson K., Francke U., Scheller R.;
 RT "Structure and chromosomal localization of the mammalian agrin gene.";
 RL J. Neurosci. 12:3535-3544(1992).
 CC -!- FUNCTION: Component of the basal lamina that causes the
 aggregation of acetylcholine receptors and acetylcholine-esterase
 on the surface of muscle fibers of the neuromuscular junction.
 CC -!- SUBUNIT: Binds to laminin.
 CC -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
 junction.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=5;
 CC Comment=Additional isoforms seem to exist. Isoforms differ in
 their acetylcholine receptor clustering activity;
 CC Name=1;
 CC IsoId=P25304-1; Sequences=Displayed;
 CC Name=2;
 CC IsoId=P25304-2; Sequences=VSP_001365;
 CC Name=3;
 CC IsoId=P25304-3; Sequences=VSP_001366;
 CC Name=4;
 CC IsoId=P25304-4; Sequences=VSP_001367;
 CC Name=5;
 CC IsoId=P25304-5; Sequences=VSP_001368;
 CC -!- TISSUE SPECIFICITY: Embryonic nervous system and muscle.
 CC -!- DEVELOPMENTAL STAGE: More abundant early in development.
 CC -!- PFM: Contains heparan sulfate chains as well as N-linked and O-
 linked oligosaccharides (By similarity).
 CC -!- SIMILARITY: Contains 9 Kazal-like domains.
 CC -!- SIMILARITY: Contains 2 laminin EGF-like domains.
 CC -!- SIMILARITY: Contains 4 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 SEA domain.
 CC -!- SIMILARITY: Contains 3 laminin G-like domains.

CC -!- CAUTION: It is uncertain whether Met-1, Met-18 or Met-24 is the
 initiator.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M64780; AAA40703.1; -;
 CC EMBL; M64780; AAA40702.1; ALT_INIT.
 CC EIR; J03039; AGRT.
 CC HSP; P00740; IEDM.
 CC InterPro; IPR008985; ConA_like_1ec_gl.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR003645; FOLN.
 CC InterPro; IPR002350; Kazal.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR001791; Laminin_G.
 CC InterPro; IPR000082; SEA_domain.
 CC Pfam; PF00008; EGF; 4.
 CC Pfam; PF00050; Kazal; 9.
 CC Pfam; PF00053; Laminin_EGF; 2.
 CC Pfam; PF00054; Laminin_G; 3.
 CC Pfam; PF01390; SEA; 1.
 CC PRINTS; PR00011; EGF_LAMININ.
 CC SMART; SM00180; EGF_Lam; 2.
 CC SMART; SM00274; FOLN; 8.
 CC SMART; SM00280; KAZAL; 9.
 CC SMART; SM00282; Lang; 3.
 CC SMART; SM00200; SEA; 1.
 CC PROSITE; PS00022; EGF_1; 6.
 CC PROSITE; PS01186; EGF_2; 1.
 CC PROSITE; PS00026; EGF_3; 4.
 CC PROSITE; PS00025; LAM_G_DOMAIN; 3.
 CC PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
 CC PROSITE; PS00024; SEA; 1.
 CC Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
 CC Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
 CC SIGNAL; 1 29
 CC CHAIN; 30 1959
 CC AGRI.
 CC DOMAIN; 65 137
 CC KAZAL-LIKE 1.
 CC KAZAL-LIKE 2.
 CC KAZAL-LIKE 3.
 CC KAZAL-LIKE 4.
 CC KAZAL-LIKE 5.
 CC KAZAL-LIKE 6.
 CC KAZAL-LIKE 7.
 CC KAZAL-LIKE 8.
 CC KAZAL-LIKE 9.
 CC LAMININ EGF-LIKE 1.
 CC LAMININ EGF-LIKE 2.
 CC LAMININ EGF-LIKE 3.
 CC LAMININ G-LIKE 1.
 CC LAMININ G-LIKE 2.
 CC LAMININ G-LIKE 3.
 CC LAMININ G-LIKE 4.
 CC LAMININ G-LIKE 5.
 CC LAMININ G-LIKE 6.
 CC LAMININ G-LIKE 7.
 CC LAMININ G-LIKE 8.
 CC LAMININ G-LIKE 9.
 CC LAMININ G-LIKE 10.
 CC LAMININ G-LIKE 11.
 CC LAMININ G-LIKE 12.
 CC LAMININ G-LIKE 13.
 CC LAMININ G-LIKE 14.
 CC LAMININ G-LIKE 15.
 CC LAMININ G-LIKE 16.
 CC LAMININ G-LIKE 17.
 CC LAMININ G-LIKE 18.
 CC LAMININ G-LIKE 19.
 CC LAMININ G-LIKE 20.
 CC LAMININ G-LIKE 21.
 CC LAMININ G-LIKE 22.
 CC LAMININ G-LIKE 23.
 CC LAMININ G-LIKE 24.
 CC LAMININ G-LIKE 25.
 CC LAMININ G-LIKE 26.
 CC LAMININ G-LIKE 27.
 CC LAMININ G-LIKE 28.
 CC LAMININ G-LIKE 29.
 CC LAMININ G-LIKE 30.
 CC LAMININ G-LIKE 31.
 CC LAMININ G-LIKE 32.
 CC LAMININ G-LIKE 33.
 CC LAMININ G-LIKE 34.
 CC LAMININ G-LIKE 35.
 CC LAMININ G-LIKE 36.
 CC LAMININ G-LIKE 37.
 CC LAMININ G-LIKE 38.
 CC LAMININ G-LIKE 39.
 CC LAMININ G-LIKE 40.
 CC LAMININ G-LIKE 41.
 CC LAMININ G-LIKE 42.
 CC LAMININ G-LIKE 43.
 CC LAMININ G-LIKE 44.
 CC LAMININ G-LIKE 45.
 CC LAMININ G-LIKE 46.
 CC LAMININ G-LIKE 47.
 CC LAMININ G-LIKE 48.
 CC LAMININ G-LIKE 49.
 CC LAMININ G-LIKE 50.
 CC LAMININ G-LIKE 51.
 CC LAMININ G-LIKE 52.
 CC LAMININ G-LIKE 53.
 CC LAMININ G-LIKE 54.
 CC LAMININ G-LIKE 55.
 CC LAMININ G-LIKE 56.
 CC LAMININ G-LIKE 57.
 CC LAMININ G-LIKE 58.
 CC LAMININ G-LIKE 59.
 CC LAMININ G-LIKE 60.
 CC LAMININ G-LIKE 61.
 CC LAMININ G-LIKE 62.
 CC LAMININ G-LIKE 63.
 CC LAMININ G-LIKE 64.
 CC LAMININ G-LIKE 65.
 CC LAMININ G-LIKE 66.
 CC LAMININ G-LIKE 67.
 CC LAMININ G-LIKE 68.
 CC LAMININ G-LIKE 69.
 CC LAMININ G-LIKE 70.
 CC LAMININ G-LIKE 71.
 CC LAMININ G-LIKE 72.
 CC LAMININ G-LIKE 73.
 CC LAMININ G-LIKE 74.
 CC LAMININ G-LIKE 75.
 CC LAMININ G-LIKE 76.
 CC LAMININ G-LIKE 77.
 CC LAMININ G-LIKE 78.
 CC LAMININ G-LIKE 79.
 CC LAMININ G-LIKE 80.
 CC LAMININ G-LIKE 81.
 CC LAMININ G-LIKE 82.
 CC LAMININ G-LIKE 83.
 CC LAMININ G-LIKE 84.
 CC LAMININ G-LIKE 85.
 CC LAMININ G-LIKE 86.
 CC LAMININ G-LIKE 87.
 CC LAMININ G-LIKE 88.
 CC LAMININ G-LIKE 89.
 CC LAMININ G-LIKE 90.
 CC LAMININ G-LIKE 91.
 CC LAMININ G-LIKE 92.
 CC LAMININ G-LIKE 93.
 CC LAMININ G-LIKE 94.
 CC LAMININ G-LIKE 95.
 CC LAMININ G-LIKE 96.
 CC LAMININ G-LIKE 97.
 CC LAMININ G-LIKE 98.
 CC LAMININ G-LIKE 99.
 CC LAMININ G-LIKE 100.
 CC LAMININ G-LIKE 101.
 CC LAMININ G-LIKE 102.
 CC LAMININ G-LIKE 103.
 CC LAMININ G-LIKE 104.
 CC LAMININ G-LIKE 105.
 CC LAMININ G-LIKE 106.
 CC LAMININ G-LIKE 107.
 CC LAMININ G-LIKE 108.
 CC LAMININ G-LIKE 109.
 CC LAMININ G-LIKE 110.
 CC LAMININ G-LIKE 111.
 CC LAMININ G-LIKE 112.
 CC LAMININ G-LIKE 113.
 CC LAMININ G-LIKE 114.
 CC LAMININ G-LIKE 115.
 CC LAMININ G-LIKE 116.
 CC LAMININ G-LIKE 117.
 CC LAMININ G-LIKE 118.
 CC LAMININ G-LIKE 119.
 CC LAMININ G-LIKE 120.
 CC LAMININ G-LIKE 121.
 CC LAMININ G-LIKE 122.
 CC LAMININ G-LIKE 123.
 CC LAMININ G-LIKE 124.
 CC LAMININ G-LIKE 125.
 CC LAMININ G-LIKE 126.
 CC LAMININ G-LIKE 127.
 CC LAMININ G-LIKE 128.
 CC LAMININ G-LIKE 129.
 CC LAMININ G-LIKE 130.
 CC LAMININ G-LIKE 131.
 CC LAMININ G-LIKE 132.
 CC LAMININ G-LIKE 133.
 CC LAMININ G-LIKE 134.
 CC LAMININ G-LIKE 135.
 CC LAMININ G-LIKE 136.
 CC LAMININ G-LIKE 137.
 CC LAMININ G-LIKE 138.
 CC LAMININ G-LIKE 139.
 CC LAMININ G-LIKE 140.
 CC LAMININ G-LIKE 141.
 CC LAMININ G-LIKE 142.
 CC LAMININ G-LIKE 143.
 CC LAMININ G-LIKE 144.
 CC LAMININ G-LIKE 145.
 CC LAMININ G-LIKE 146.
 CC LAMININ G-LIKE 147.
 CC LAMININ G-LIKE 148.
 CC LAMININ G-LIKE 149.
 CC LAMININ G-LIKE 150.
 CC LAMININ G-LIKE 151.
 CC LAMININ G-LIKE 152.
 CC LAMININ G-LIKE 153.
 CC LAMININ G-LIKE 154.
 CC LAMININ G-LIKE 155.
 CC LAMININ G-LIKE 156.
 CC LAMININ G-LIKE 157.
 CC LAMININ G-LIKE 158.
 CC LAMININ G-LIKE 159.
 CC LAMININ G-LIKE 160.
 CC LAMININ G-LIKE 161.
 CC LAMININ G-LIKE 162.
 CC LAMININ G-LIKE 163.
 CC LAMININ G-LIKE 164.
 CC LAMININ G-LIKE 165.
 CC LAMININ G-LIKE 166.
 CC LAMININ G-LIKE 167.
 CC LAMININ G-LIKE 168.
 CC LAMININ G-LIKE 169.
 CC LAMININ G-LIKE 170.
 CC LAMININ G-LIKE 171.
 CC LAMININ G-LIKE 172.
 CC LAMININ G-LIKE 173.
 CC LAMININ G-LIKE 174.
 CC LAMININ G-LIKE 175.
 CC LAMININ G-LIKE 176.
 CC LAMININ G-LIKE 177.
 CC LAMININ G-LIKE 178.
 CC LAMININ G-LIKE 179.
 CC LAMININ G-LIKE 180.
 CC LAMININ G-LIKE 181.
 CC LAMININ G-LIKE 182.
 CC LAMININ G-LIKE 183.
 CC LAMININ G-LIKE 184.
 CC LAMININ G-LIKE 185.
 CC LAMININ G-LIKE 186.
 CC LAMININ G-LIKE 187.
 CC LAMININ G-LIKE 188.
 CC LAMININ G-LIKE 189.
 CC LAMININ G-LIKE 190.
 CC LAMININ G-LIKE 191.
 CC LAMININ G-LIKE 192.
 CC LAMININ G-LIKE 193.
 CC LAMININ G-LIKE 194.
 CC LAMININ G-LIKE 195.
 CC LAMININ G-LIKE 196.
 CC LAMININ G-LIKE 197.
 CC LAMININ G-LIKE 198.
 CC LAMININ G-LIKE 199.
 CC LAMININ G-LIKE 200.
 CC LAMININ G-LIKE 201.
 CC LAMININ G-LIKE 202.
 CC LAMININ G-LIKE 203.
 CC LAMININ G-LIKE 204.
 CC LAMININ G-LIKE 205.
 CC LAMININ G-LIKE 206.
 CC LAMININ G-LIKE 207.
 CC LAMININ G-LIKE 208.
 CC LAMININ G-LIKE 209.
 CC LAMININ G-LIKE 210.
 CC LAMININ G-LIKE 211.
 CC LAMININ G-LIKE 212.
 CC LAMININ G-LIKE 213.
 CC LAMININ G-LIKE 214.
 CC LAMININ G-LIKE 215.
 CC LAMININ G-LIKE 216.
 CC LAMININ G-LIKE 217.
 CC LAMININ G-LIKE 218.
 CC LAMININ G-LIKE 219.
 CC LAMININ G-LIKE 220.
 CC LAMININ G-LIKE 221.
 CC LAMININ G-LIKE 222.
 CC LAMININ G-LIKE 223.
 CC LAMININ G-LIKE 224.
 CC LAMININ G-LIKE 225.
 CC LAMININ G-LIKE 226.
 CC LAMININ G-LIKE 227.
 CC LAMININ G-LIKE 228.
 CC LAMININ G-LIKE 229.
 CC LAMININ G-LIKE 230.
 CC LAMININ G-LIKE 231.
 CC LAMININ G-LIKE 232.
 CC LAMININ G-LIKE 233.
 CC LAMININ G-LIKE 234.
 CC LAMININ G-LIKE 235.
 CC LAMININ G-LIKE 236.
 CC LAMININ G-LIKE 237.
 CC LAMININ G-LIKE 238.
 CC LAMININ G-LIKE 239.
 CC LAMININ G-LIKE 240.
 CC LAMININ G-LIKE 241.
 CC LAMININ G-LIKE 242.
 CC LAMININ G-LIKE 243.
 CC LAMININ G-LIKE 244.
 CC LAMININ G-LIKE 245.
 CC LAMININ G-LIKE 246.
 CC LAMININ G-LIKE 247.
 CC LAMININ G-LIKE 248.
 CC LAMININ G-LIKE 249.
 CC LAMININ G-LIKE 250.
 CC LAMININ G-LIKE 251.
 CC LAMININ G-LIKE 252.
 CC LAMININ G-LIKE 253.
 CC LAMININ G-LIKE 254.
 CC LAMININ G-LIKE 255.
 CC LAMININ G-LIKE 256.
 CC LAMININ G-LIKE 257.
 CC LAMININ G-LIKE 258.
 CC LAMININ G-LIKE 259.
 CC LAMININ G-LIKE 260.
 CC LAMININ G-LIKE 261.
 CC LAMININ G-LIKE 262.
 CC LAMININ G-LIKE 263.
 CC LAMININ G-LIKE 264.
 CC LAMININ G-LIKE 265.
 CC LAMININ G-LIKE 266.
 CC LAMININ G-LIKE 267.
 CC LAMININ G-LIKE 268.
 CC LAMININ G-LIKE 269.
 CC LAMININ G-LIKE 270.
 CC LAMININ G-LIKE 271.
 CC LAMININ G-LIKE 272.
 CC LAMININ G-LIKE 273.
 CC LAMININ G-LIKE 274.
 CC LAMININ G-LIKE 275.
 CC LAMININ G-LIKE 276.
 CC LAMININ G-LIKE 277.
 CC LAMININ G-LIKE 278.
 CC LAMININ G-LIKE 279.
 CC LAMININ G-LIKE 280.
 CC LAMININ G-LIKE 281.
 CC LAMININ G-LIKE 282.
 CC LAMININ G-LIKE 283.
 CC LAMININ G-LIKE 284.
 CC LAMININ G-LIKE 285.
 CC LAMININ G-LIKE 286.
 CC LAMININ G-LIKE 287.
 CC LAMININ G-LIKE 288.
 CC LAMININ G-LIKE 289.
 CC LAMININ G-LIKE 290.
 CC LAMININ G-LIKE 291.
 CC LAMININ G-LIKE 292.
 CC LAMININ G-LIKE 293.
 CC LAMININ G-LIKE 294.
 CC LAMININ G-LIKE 295.
 CC LAMININ G-LIKE 296.
 CC LAMININ G-LIKE 297.
 CC LAMININ G-LIKE 298.
 CC LAMININ G-LIKE 299.
 CC LAMININ G-LIKE 300.
 CC LAMININ G-LIKE 301.
 CC LAMININ G-LIKE 302.
 CC LAMININ G-LIKE 303.
 CC LAMININ G-LIKE 304.
 CC LAMININ G-LIKE 305.
 CC LAMININ G-LIKE 306.
 CC LAMININ G-LIKE 307.
 CC LAMININ G-LIKE 308.
 CC LAMININ G-LIKE 309.
 CC LAMININ G-LIKE 310.
 CC LAMININ G-LIKE 311.
 CC LAMININ G-LIKE 312.
 CC LAMININ G-LIKE 313.
 CC LAMININ G-LIKE 314.
 CC LAMININ G-LIKE 315.
 CC LAMININ G-LIKE 316.
 CC LAMININ G-LIKE 317.
 CC LAMININ G-LIKE 318.
 CC LAMININ G-LIKE 319.
 CC LAMININ G-LIKE 320.
 CC LAMININ G-LIKE 321.
 CC LAMININ G-LIKE 322.
 CC LAMININ G-LIKE 323.
 CC LAMININ G-LIKE 324.
 CC LAMININ G-LIKE 325.
 CC LAMININ G-LIKE 326.
 CC LAMININ G-LIKE 327.
 CC LAMININ G-LIKE 328.
 CC LAMININ G-LIKE 329.
 CC LAMININ G-LIKE 330.
 CC LAMININ G-LIKE 331.
 CC LAMININ G-LIKE 332.
 CC LAMININ G-LIKE 333.
 CC LAMININ G-LIKE 334.
 CC LAMININ G-LIKE 335.
 CC LAMININ G-LIKE 336.
 CC LAMININ G-LIKE 337.
 CC LAMININ G-LIKE 338.
 CC LAMININ G-LIKE 339.
 CC LAMININ G-LIKE 340.
 CC LAMININ G-LIKE 341.
 CC LAMININ G-LIKE 342.
 CC LAMININ G-LIKE 343.
 CC LAMININ G-LIKE 344.
 CC LAMININ G-LIKE 345.
 CC LAMININ G-LIKE 346.
 CC LAMININ G-LIKE 347.
 CC LAMININ G-LIKE 348.
 CC LAMININ G-LIKE 349.
 CC LAMININ G-LIKE 350.
 CC LAMININ G-LIKE 351.
 CC LAMININ G-LIKE 352.
 CC LAMININ G-LIKE 353.
 CC LAMININ G-LIKE 354.
 CC LAMININ G-LIKE 355.
 CC LAMININ G-LIKE 356.
 CC LAMININ G-LIKE 357.
 CC LAMININ G-LIKE 358.
 CC LAMININ G-LIKE 359.
 CC LAMININ G-LIKE 360.
 CC LAMININ G-LIKE 361.
 CC LAMININ G-LIKE 362.
 CC LAMININ G-LIKE 363.
 CC LAMININ G-LIKE 364.
 CC LAMININ G-LIKE 365.
 CC LAMININ G-LIKE 366.
 CC LAMININ G-LIKE 367.
 CC LAMININ G-LIKE 368.
 CC LAMININ G-LIKE 369.
 CC LAMININ G-LIKE 370.
 CC LAMININ G-LIKE 371.
 CC LAMININ G-LIKE 372.
 CC LAMININ G-LIKE 373.
 CC LAMININ G-LIKE 374.
 CC LAMININ G-LIKE 375.
 CC LAMININ G-LIKE 376.
 CC LAMININ G-LIKE 377.
 CC LAMININ G-LIKE 378.
 CC LAMININ G-LIKE 379.
 CC LAMININ G-LIKE 380.
 CC LAMININ G-LIKE 381.
 CC LAMININ G-LIKE 382.
 CC LAMININ G-LIKE 383.
 CC LAMININ G-LIKE 384.
 CC LAMININ G-LIKE 385.
 CC LAMININ G-LIKE 386.
 CC LAMININ G-LIKE 387.
 CC LAMININ G-LIKE 388.
 CC LAMININ G-LIKE 389.
 CC LAMININ G-LIKE 390.
 CC LAMININ G-LIKE 391.
 CC LAMININ G-LIKE 392.
 CC LAMININ G-LIKE 393.
 CC LAMININ G-LIKE 394.
 CC LAMININ G-LIKE 395.
 CC LAMININ G-LIKE 396.
 CC LAMININ G-LIKE 397.
 CC LAMININ G-LIKE 398.
 CC LAMININ G-LIKE 399.
 CC LAMININ G-LIKE 400.
 CC LAMININ G-LIKE 401.
 CC LAMININ G-LIKE 402.
 CC LAMININ G-LIKE 403.
 CC LAMININ G-LIKE 404.
 CC LAMININ G-LIKE 405.
 CC LAMININ G-LIKE 406.
 CC LAMININ G-LIKE 407.
 CC LAMININ G-LIKE 408.
 CC LAMININ G-LIKE 409.
 CC LAMININ G-LIKE 410.
 CC LAMININ G-LIKE 411.
 CC LAMININ G-LIKE 412.
 CC LAMININ G-LIKE 413.
 CC LAMININ G-LIKE 414.
 CC LAMININ G-LIKE 415.
 CC LAMININ G-LIKE 416.
 CC LAMININ G-LIKE 417.
 CC LAMININ G-LIKE 418.
 CC LAMININ G-LIKE 419.
 CC LAMININ G-LIKE 420.
 CC LAMININ G-LIKE 421.
 CC LAMININ G-LIKE 422.
 CC LAMININ G-LIKE 423.
 CC LAMININ G-LIKE 424.
 CC LAMININ G-LIKE 425.
 CC LAMININ G-LIKE 426.
 CC LAMININ G-LIKE 427.
 CC LAMININ G-LIKE 428.
 CC LAMININ G-LIKE 429.
 CC LAMININ G-LIKE 430.
 CC LAMININ G-LIKE 431.
 CC LAMININ G-LIKE 432.
 CC LAMININ G-LIKE 433.
 CC LAMININ G-LIKE 434.
 CC LAMININ G-LIKE 435.
 CC LAMININ G-LIKE 436.
 CC LAMININ G-LIKE 437.
 CC LAMININ G-LIKE 438.
 CC LAMININ G-LIKE 439.
 CC LAMININ G-LIKE 440.
 CC LAMININ G-LIKE 441.
 CC LAMININ G-LIKE 442.
 CC LAMININ G-LIKE 443.
 CC LAMININ G-LIKE 444.
 CC LAMININ G-LIKE 445.
 CC LAMININ G-LIKE 446.
 CC LAMININ G-LIKE 447.
 CC LAMININ G-LIKE 448.
 CC LAMININ G-LIKE 449.
 CC LAMININ G-LIKE 450.
 CC LAMININ G-LIKE 451.
 CC LAMININ G-LIKE 452.
 CC LAMININ G-LIKE 453.
 CC LAMININ G-LIKE 454.
 CC LAMININ G-LIKE 455.
 CC LAMININ G-LIKE 456.
 CC LAMININ G-LIKE 457.
 CC LAMININ G-LIKE 458.
 CC LAMININ G-LIKE 459.
 CC LAMININ G-LIKE 460.
 CC LAMININ G-LIKE 461.
 CC LAMININ G-LIKE 462.
 CC LAMININ G-LIKE 463.
 CC LAMININ G-LIKE 464.
 CC LAMININ G-LIKE 465.
 CC LAMININ G-LIKE 466.
 CC LAMININ G-LIKE 467.
 CC LAMININ G-LIKE 468.
 CC LAMININ G-LIKE 469.
 CC LAMININ G-LIKE 470.
 CC LAMININ G-LIKE 471.
 CC LAMININ G-LIKE 472.
 CC LAMININ G-LIKE 473.
 CC LAMININ G-LIKE 474.
 CC LAMININ G-LIKE 475.
 CC LAMININ G-LIKE 476.
 CC LAMININ G-LIKE 477.
 CC LAMININ G-LIKE 478.
 CC LAMININ G-LIKE 479.
 CC LAMININ G-LIKE 480.
 CC LAMININ G-LIKE 481.
 CC LAMININ G-LIKE 482.
 CC LAMININ G-LIKE 483.
 CC LAMININ G-LIKE 484.
 CC LAMININ G-LIKE 485.
 CC LAMININ G-LIKE 486.
 CC LAMININ G-LIKE 487.
 CC LAMININ G-LIKE 488.
 CC LAMININ G-LIKE 489.
 CC LAMININ G-LIKE 490.
 CC LAMININ G-LIKE 491.
 CC LAMININ G-LIKE 492.
 CC LAMININ G-LIKE 493.
 CC LAMININ G-LIKE 494.
 CC LAMININ G-LIKE 495.
 CC LAMININ G-LIKE 496.
 CC LAMININ G-LIKE 497.
 CC LAMININ G-LIKE 498.
 CC LAMININ G-LIKE 499.
 CC LAMININ G-LIKE 500.
 CC LAMININ G-LIKE 501.
 CC LAMININ G-LIKE 502.
 CC LAMININ G-LIKE 503.
 CC LAMININ G-LIKE 504.
 CC LAMININ G-LIKE 505.
 CC LAMININ G-LIKE 506.
 CC LAMININ G-LIKE 507.
 CC LAMININ G-LIKE 508.
 CC LAMININ G-LIKE 509.
 CC LAMININ G-LIKE 510.
 CC LAMININ G-LIKE 511.
 CC LAMININ G-LIKE 512.
 CC LAMININ G-LIKE 513.
 CC LAMININ G-LIKE 514.
 CC LAMININ G-LIKE 515.
 CC LAMININ G-LIKE 516.
 CC LAMININ G-LIKE 517.
 CC LAMININ G-LIKE 518.
 CC LAMININ G-LIKE 519.
 CC LAMININ G-LIKE 520.
 CC LAMININ G-LIKE 521.
 CC LAMININ G-LIKE 522.
 CC LAMININ G-LIKE 523.
 CC LAMININ G-LIKE 524.
 CC LAMININ G-LIKE 525.
 CC LAMININ G-LIKE 526.
 CC LAMININ G-LIKE 527.
 CC LAMININ G-LIKE 528.
 CC LAMININ G-LIKE 529.
 CC LAMININ G-LIKE 530.
 CC LAMININ G-LIKE 531.
 CC LAMININ G-LIKE 532.
 CC LAMININ G-LIKE 533.
 CC LAMININ G-LIKE 534.
 CC LAMININ G-LIKE 535.
 CC LAMININ G-LIKE 536.
 CC LAMININ G-LIKE 537.
 CC LAMININ G-LIKE 538.
 CC LAMININ G-LIKE 539.
 CC LAMININ G-LIKE 540.
 CC LAMININ G-LIKE 541.
 CC LAMININ G-LIKE 542.
 CC LAMININ G-LIKE 543.
 CC LAMININ G-LIKE 544.
 CC LAMININ G-LIKE 545.
 CC LAMININ G-LIKE 546.
 CC LAMININ G-LIKE 547.
 CC LAMININ G-LIKE 548.
 CC LAMININ G-LIKE 549.
 CC LAMININ G-LIKE 550.
 CC LAMININ G-LIKE 551.
 CC LAMININ G-LIKE 552.
 CC LAMININ G-LIKE 553.
 CC LAMININ G-LIKE 554.
 CC LAMININ G-LIKE 555.
 CC LAMININ G-LIKE 556.
 CC LAMININ G-LIKE 557.
 CC LAMININ G-LIKE 558.
 CC LAMININ G-LIKE 559.
 CC LAMININ G-LIKE 560.
 CC LAMININ G-LIKE 561.
 CC LAMININ G-LIKE 562.
 CC LAMININ G-LIKE 563.
 CC LAMININ G-LIKE 564.
 CC LAMININ G-LIKE 565.
 CC LAMININ G-LIKE 566.
 CC LAMININ G-LIKE 567.
 CC LAMININ G-LIKE 568.
 CC LAMININ G-LIKE 569.
 CC LAMININ G-LIKE 570.
 CC LAMININ G-LIKE 571.
 CC LAMININ G-LIKE 572.
 CC LAMININ G-LIKE 573.
 CC LAMININ G-LIKE 574.
 CC LAMININ G-LIKE 575.

FT DISULFID 389 408 POTENTIAL.
 FT DISULFID 397 429 POTENTIAL.
 FT DISULFID 454 473 POTENTIAL.
 FT DISULFID 462 494 POTENTIAL.
 FT DISULFID 518 538 POTENTIAL.
 FT DISULFID 527 559 POTENTIAL.
 FT DISULFID 604 624 POTENTIAL.
 FT DISULFID 613 645 POTENTIAL.
 FT DISULFID 688 707 BY SIMILARITY.
 FT DISULFID 690 700 BY SIMILARITY.
 FT DISULFID 709 718 BY SIMILARITY.
 FT DISULFID 721 739 BY SIMILARITY.
 FT DISULFID 742 754 BY SIMILARITY.
 FT DISULFID 744 761 BY SIMILARITY.
 FT DISULFID 763 772 BY SIMILARITY.
 FT DISULFID 775 786 BY SIMILARITY.
 FT DISULFID 823 843 POTENTIAL.
 FT DISULFID 832 864 POTENTIAL.
 FT DISULFID 1224 1235 BY SIMILARITY.
 FT DISULFID 1229 1246 BY SIMILARITY.
 FT DISULFID 1248 1257 BY SIMILARITY.
 FT DISULFID 1444 1455 POTENTIAL.
 FT DISULFID 1467 1476 POTENTIAL.
 FT DISULFID 1483 1494 BY SIMILARITY.
 FT DISULFID 1488 1504 BY SIMILARITY.
 FT DISULFID 1506 1515 BY SIMILARITY.
 FT DISULFID 1713 1727 BY SIMILARITY.
 FT DISULFID 1721 1736 BY SIMILARITY.
 FT DISULFID 1738 1747 BY SIMILARITY.
 FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 672 672 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 957 957 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 1144 1152 Missing (in isoform 2).
 FT VARSPLIC 1780 1798 Missing (in isoform 3).
 FT VARSPLIC 1788 1798 Missing (in isoform 4).
 FT VARSPLIC 1780 1787 Missing (in isoform 5).
 FT VARIANT 314 314 V -> VTCD (IN A VARIANT).
 SQ SEQUENCE 1959 AA; 208645 MW; 7FEFDFDAFF89CC31 CRC64;

Query Match 22.5%; Score 859; DB 1; Length 1959;
 Best Local Similarity 31.5%; Pred No 1.4e-46;
 Matches 241; Conservative 92; Mismatches 278; Indels 154; Gaps 23;

Qy 2 IKITFRPDSADGMLLYNQKRVPGSPNTLANRQPDFISGLVGGPFRFPDAGSGMATIR 61
 Db LALEFRALETEGLLYNGNA-----RQKDFLALALLDGRVQFRFDTGSGPAVLT 1335

Qy 62 HPTPLALGHFTVTLRLSLTQSLIVGDLAPVNGTSQKFGQLDLNLESLYLGYPDYG-- 119
 Db SLVPEVFGHRHLESLSHWEGTSLVDGETPVVGESPSCTDGLNDLTNLYVGGIPBEQVA 1395

Qy 120 -AIPAGLSGFGTCTVRELRIQEEIVFHDNLNTA---HGSHC--PTCRDRPCQNGQC 173
 Db MVLDRTSVGLKGCINMLDINNQQLESDQRAAVQSSGVGECGDHPCLNPNCHGGALC 1455

Qy 174 HDSRSSYVVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRVTCRCHLGRSL 233
 Db QALEAGMFLCQCPGRGFGPTCADEKS-PCQNPCHGAAPCVLSSG-GAKCECLGRSGT 1513

Qy 234 RCEGVTVTPFSLSGAG-----SYLALPALTNTNTHL-----RLDVEFKPLADPG 278
 Db FCQ---TV---LETAGSRPFLADFNFGSYLELKGHTFERDLGKMALEMYFLARGPSG 1566

Qy 279 VLFPSGKGFVEDFVSLAMVVGHLFPRYELSGSLAVLSASPLALGRWHRVSAERLNDK 338
 Db LLLYNGQKTDGKGFVSLALHNEHLEFCYDLGLGKAIVRSKEPIALGTWVFLERNGRK 1626

Qy 339 GSLRVNGGRPVLRSSPGKSGQ-----LNLHTLLYLGGVEPSVPLSPATNMSAHPRGCVGE 393
 Db 1627 GALQVCDGPERVLGESP-KSRKVPHTMLNKEPLYIGGAPDFSKLARGAAYSSGSGVIQL 1685

Qy 394 VSVNGKRLDLYSFLGSGIGQCTYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFPGKDL 453
 Db 1686 VSLRGHQL-----LTQEHVLRADVSPFADHPCTQA----- 1716

Qy 454 CEHEENPCQLREFPCLHGTC---QOT-RLCLPLGFSQPCQCGSGHGFIARSDWHLESGG 509
 Db 1717 -----LGNPCLNGGSCVPERATVECLCPGFGSLGHCXGL-----VKSXVG 1757

Qy 510 NDAPGOYGAFFHDDGFLAPGHVFSRSLPEV-----PETI----- 544
 Db 1758 -----DLETPLADGRTYEYLNVAIEBELTNEIPAPETLDSRALFSKALQSN 1805

Qy 545 --ELEVRTSTASGLLLWQGVGEAGQKQKFIISLQDQHLVFRYQLGSGEARLVSEDP 602
 Db 1806 HFELSRLTTEATQGLVLM-----IGKAAERADYNALAIVDGHLQLSYDLGSPVVLRSVKV 1861

Qy 603 NDGEWHRTALREGRGSGSIQVDGELVSGRSPGNVAVNAKGSVYVIGAP--DVATLTGG 660
 Db 1862 NTNRLRTPRAHREHREGSLQVNEAPVTGSSPLGATQDITDGLWLGLQKLPVGOALPK 1921

Qy 661 RFSSGITGCVKMLVLSARPGAPPQPLDLQFRAQAGANTRPCPS 705
 Db 1922 AYCTGPFVGLRDVVVGH-----ROLHLEDAVTKPELAPCPT 1958

RESULT 4
 AGRI_CHICK STANDARD; PRT; 1955 AA.
 ID AGRI_CHICK
 AC P31696;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Agrin precursor.
 GN AGRI.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=92232297; PubMed=1314620;
 RA Tsai M.A., Ruegg M.A., Escher G., Kroeger S., McMahan U.J.;
 RT "cDNA that encodes active agrin.";
 RL Neuron 8:677-689(1992).
 RN [2]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=92232298; PubMed=1314621;
 RA Ruegg M.A., Tsai M.A., Horton S.E., Kroeger S., Escher G.,
 RA Gensch E.M., McMahan U.J.;
 RT "The agrin gene codes for a family of basal lamina proteins that
 differ in function and distribution."
 RL Neuron 8:691-699(1992).
 CC -1- FUNCTION: Component of the basal lamina that causes the
 aggregation of acetylcholine receptors and acetylcholine-esterase
 on the surface of muscle fibers of the neuromuscular junction.
 CC -1- SUBUNIT: Binds to laminin.
 CC -1- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
 junction.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Comment=Additional isoforms seem to exist. Isoforms differ in
 their acetylcholine receptor clustering activity;
 CC Name=1;
 CC IsoId=P31696-1; Sequence=Displayed;
 CC Name=2; Synonyms=Agrin-related protein 1;
 CC IsoId=P31696-2; Sequence=VSP_001370;
 CC Name=3; Synonyms=Agrin-related protein 2;

InterPro; IPR003599; Ig. 202
 InterPro; IPR003598; Ig C2. 190
 InterPro; IPR00034; Laminin B. 215
 InterPro; IPR002049; Laminin_EGF. 224
 InterPro; IPR001791; Laminin_G. 226
 InterPro; IPR002172; LDL_receptor_A. 240
 Pfam; PF00047; Ig; 16. 259
 Pfam; PF00052; laminin B; 2. 268
 Pfam; PF00053; laminin EGF; 5. 955
 Pfam; PF00057; ldl receptor a; 3. 964
 PRINTS; PR00261; LDLRECEPTOR. 971
 PRINTS; PR00261; LDLRECEPTOR. 986
 ProDom; PD003031; Laminin_B; 2. 1002
 SMART; SM00181; EGF; 5. 1011
 SMART; SM00179; EGF_CA; 2. 1021
 SMART; SM00180; EGF_Lam; 6. 1027
 SMART; SM00409; IG; 17. 1030
 SMART; SM00408; IGC2; 17. 1039
 SMART; SM00281; Lam; 1. 1042
 SMART; SM00282; Lam; 3. 1058
 SMART; SM00192; LDLa; 3. 1061
 SMART; SM0022; EGF; 1; 7. 1079
 PROSITE; PS00022; EGF_1; 7. 1082
 PROSITE; PS01185; EGF_2; 2. 1094
 PROSITE; PS00025; LAM_G DOMAIN; 3. 1152
 PROSITE; PS01248; LAMININ_TYPE_EGF; 7. 1338
 PROSITE; PS01209; LDLRA 1; 3. 1435
 PROSITE; PS00068; LDLRA 2; 3. 1527
 Proteoglycan; Glycoprotein; Signal; Immunoglobulin domain; Repeat; 1683
 Basement membrane; Extracellular matrix; Alternative splicing; 1719
 Laminin EGF-like domain. 1814
 SIGNAL 22 POTENTIAL. 1861
 CHAIN 23 3375 BASEMENT MEMBRANE PROTEOGLYCAN. 1907
 DOMAIN 45 130 IG-LIKE C2-TYPE 1. 1954
 DOMAIN 148 184 LDL-RECEPTOR CLASS A 1. 2039
 DOMAIN 189 225 LDL-RECEPTOR CLASS A 2. 2147
 DOMAIN 232 269 LDL-RECEPTOR CLASS A 3. 2242
 DOMAIN 271 355 IG-LIKE C2-TYPE 2. 2329
 DOMAIN 384 431 LAMININ EGF-LIKE 1 (INCOMPLETE). 2420
 DOMAIN 432 441 LAMININ EGF-LIKE 2 (N-TERMINAL). 2467
 DOMAIN 442 633 LAMININ EGF-LIKE 3 (INCOMPLETE). 2514
 DOMAIN 634 666 LAMININ EGF-LIKE 4 (C-TERMINAL). 2544
 DOMAIN 674 720 LAMININ EGF-LIKE 5 (INCOMPLETE). 2557
 DOMAIN 721 730 LAMININ EGF-LIKE 6 (N-TERMINAL). 2557
 DOMAIN 731 921 LAMININ EGF-LIKE 7 (INCOMPLETE). 2557
 DOMAIN 922 954 LAMININ EGF-LIKE 8 (C-TERMINAL). 2557
 DOMAIN 955 1004 LAMININ EGF-LIKE 9 (C-TERMINAL). 2557
 DOMAIN 1011 1060 LAMININ EGF-LIKE 10 (C-TERMINAL). 2557
 DOMAIN 1061 1111 LAMININ EGF-LIKE 11 (C-TERMINAL). 2557
 DOMAIN 1126 1222 LAMININ EGF-LIKE 12 (C-TERMINAL). 2557
 DOMAIN 1226 1311 LAMININ EGF-LIKE 13 (C-TERMINAL). 2557
 DOMAIN 1319 1401 LAMININ EGF-LIKE 14 (C-TERMINAL). 2557
 DOMAIN 1401 1499 LAMININ EGF-LIKE 15 (C-TERMINAL). 2557
 DOMAIN 1503 1585 LAMININ EGF-LIKE 16 (C-TERMINAL). 2557
 DOMAIN 1588 1680 LAMININ EGF-LIKE 17 (C-TERMINAL). 2557
 DOMAIN 1690 1785 LAMININ EGF-LIKE 18 (C-TERMINAL). 2557
 DOMAIN 1793 1878 LAMININ EGF-LIKE 19 (C-TERMINAL). 2557
 DOMAIN 1886 1970 LAMININ EGF-LIKE 20 (C-TERMINAL). 2557
 DOMAIN 1973 2069 LAMININ EGF-LIKE 21 (C-TERMINAL). 2557
 DOMAIN 2073 2163 LAMININ EGF-LIKE 22 (C-TERMINAL). 2557
 DOMAIN 2173 2260 LAMININ EGF-LIKE 23 (C-TERMINAL). 2557
 DOMAIN 2263 2343 LAMININ EGF-LIKE 24 (C-TERMINAL). 2557
 DOMAIN 2349 2435 LAMININ EGF-LIKE 25 (C-TERMINAL). 2557
 DOMAIN 2446 2530 LAMININ EGF-LIKE 26 (C-TERMINAL). 2557
 DOMAIN 2532 2613 LAMININ EGF-LIKE 27 (C-TERMINAL). 2557
 DOMAIN 2613 2690 LAMININ EGF-LIKE 28 (C-TERMINAL). 2557
 DOMAIN 2691 3093 LAMININ EGF-LIKE 29 (C-TERMINAL). 2557
 DOMAIN 2972 3066 LAMININ EGF-LIKE 30 (C-TERMINAL). 2557
 DOMAIN 3180 3359 LAMININ EGF-LIKE 31 (C-TERMINAL). 2557
 DISULFID 66 114 BY SIMILARITY. 2557
 DISULFID 149 161 BY SIMILARITY. 2557
 DISULFID 156 174 BY SIMILARITY. 2557
 DISULFID 168 183 BY SIMILARITY. 2557

FT DISULFID 190 202 BY SIMILARITY.
 FT DISULFID 197 215 BY SIMILARITY.
 FT DISULFID 209 224 BY SIMILARITY.
 FT DISULFID 233 246 BY SIMILARITY.
 FT DISULFID 240 259 BY SIMILARITY.
 FT DISULFID 253 268 BY SIMILARITY.
 FT DISULFID 955 964 BY SIMILARITY.
 FT DISULFID 971 986 BY SIMILARITY.
 FT DISULFID 974 986 BY SIMILARITY.
 FT DISULFID 1011 1021 BY SIMILARITY.
 FT DISULFID 1013 1027 BY SIMILARITY.
 FT DISULFID 1030 1039 BY SIMILARITY.
 FT DISULFID 1042 1058 BY SIMILARITY.
 FT DISULFID 1061 1079 BY SIMILARITY.
 FT DISULFID 1082 1094 BY SIMILARITY.
 FT DISULFID 1094 1109 BY SIMILARITY.
 FT DISULFID 1152 1200 BY SIMILARITY.
 FT DISULFID 1338 1384 BY SIMILARITY.
 FT DISULFID 1435 1481 BY SIMILARITY.
 FT DISULFID 1527 1573 BY SIMILARITY.
 FT DISULFID 1618 1663 BY SIMILARITY.
 FT DISULFID 1719 1767 BY SIMILARITY.
 FT DISULFID 1814 1861 BY SIMILARITY.
 FT DISULFID 1907 1954 BY SIMILARITY.
 FT DISULFID 1954 2039 BY SIMILARITY.
 FT DISULFID 2039 2147 BY SIMILARITY.
 FT DISULFID 2147 2242 BY SIMILARITY.
 FT DISULFID 2242 2329 BY SIMILARITY.
 FT DISULFID 2329 2420 BY SIMILARITY.
 FT DISULFID 2420 2514 BY SIMILARITY.
 FT CARBOHYD 2467 2557 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2476 2557 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2557 2557 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 3143 3143 N-LINKED (GLCNAC. .) (POTENTIAL).
 Query Match 13.4%; Score 511; DB 1; Length 3375;
 Best Local Similarity 20.9%; Pred No. 2,6e-24;
 Matches 179; Conservative 140; Mismatches 332; Indels 204; Gaps 29;
 QY 1 EKIITRPSADGMLLYNQKRV-----GPTMLANRQPDFISGLVGRPERFPAGS 55
 Db 2557 DIEITLNTANPKGIIFET--KRINSGDLATPYDTHIEAKITDYGTV---LYEFDIGN 2610
 QY 56 GNATIRHPPLALGHFTVTLRLSLTQGLSLVGLA-----PVNGTSQKFGQLDLN 107
 Db 2611 GRQIVETNPINPENNVIKNDKNQVTLQDNDSATIRQHTNPLSLSTG-----VN 2664
 QY 108 RELYLGYPDYGAIPKAGLSGFGICVRELRIOGEEIVFHDNLNTARGISHCPTCRDRPC 167
 Db 2665 RFVFIGGRHE-----PTNEANDFRGIISQVVLGHVNLGDLARIPSSVVKYDACASTNLC 2719
 QY 168 QNGGQCHDSSES--SYVVCVPAGFTGSRCHESQALHCHPEACGPDATCVNRPDGRGYTCRC 226
 Db 2720 LNCANCRANNHHGSCBCEAFHGEYQW--RNSCHDESCN-TGICLDNEE--SWQVCV 2775
 QY 227 HIGRSLRCEEGVTTTP-SLSGAGSYLALPALNTHTHELRLDVEFKPLADPG---VLLP 282
 Db 2776 PLGTTLGRCEKTEIPQLGFTSDTSFLAVKRPVKFE---SIRMKLRPOADSDSHILMYF 2832
 QY 283 SGKSGPVEDFVSLAMVGHLEFRYELGSLAVLSRAEPLALGRWHVRSAPRLNKDGLSLR 342
 Db 2833 ASDYGSNTKQVTSLSIANQVV-----LTVRRPDKVEQKIRSETLEAGEL-IDVAVR 2883
 QY 343 VNGGRVPLSSPKSQGLNLHTL-----LYLGGVEPSPVPLSPATNMAHRRGCYGVSVN 397
 Db 2884 QAGNALVMTVDGNQVSTIETDTLKPGETIFIGGLPPGLN-SFDDVVEQSFGCGVYELIN 2942
 QY 398 GKRLDITYSFLGSQG-IQCCYDS-----SPC 422
 Db 2943 SQDVLDL--QNLSSGDISSCESQPFVEEDDTTITTTTTEPEAVIEPTTTEPTTTEPI 3000

FT DOMAIN 1506 1710 LAMININ G-LIKE 1.
 FT DOMAIN 1713 1749 EGF-LIKE 4, CALCIUM-BINDING.
 FT DOMAIN 1753 1935 LAMININ G-LIKE 2.
 FT DOMAIN 1937 1972 EGF-LIKE 5, CALCIUM-BINDING.
 FT DOMAIN 1973 2011 EGF-LIKE 6, CALCIUM-BINDING.
 FT DOMAIN 2012 2044 EGF-LIKE 7, CALCIUM-BINDING.
 FT DOMAIN 2046 2081 EGF-LIKE 8, CALCIUM-BINDING.
 FT DOMAIN 2087 2120 LAMININ EGF-LIKE.
 FT DOMAIN 2120 2527 GFS.
 FT DISULFID 1370 1381 BY SIMILARITY.
 FT DISULFID 1375 1412 BY SIMILARITY.
 FT DISULFID 1414 1423 BY SIMILARITY.
 FT DISULFID 1430 1441 BY SIMILARITY.
 FT DISULFID 1435 1450 BY SIMILARITY.
 FT DISULFID 1452 1461 BY SIMILARITY.
 FT DISULFID 1470 1481 BY SIMILARITY.
 FT DISULFID 1475 1491 BY SIMILARITY.
 FT DISULFID 1493 1504 BY SIMILARITY.
 FT DISULFID 1717 1728 BY SIMILARITY.
 FT DISULFID 1722 1737 BY SIMILARITY.
 FT DISULFID 1739 1748 BY SIMILARITY.
 FT DISULFID 1941 1952 BY SIMILARITY.
 FT DISULFID 1946 1961 BY SIMILARITY.
 FT DISULFID 1963 1972 BY SIMILARITY.
 FT DISULFID 1976 1987 BY SIMILARITY.
 FT DISULFID 1981 1999 BY SIMILARITY.
 FT DISULFID 2001 2010 BY SIMILARITY.
 FT DISULFID 2018 2031 BY SIMILARITY.
 FT DISULFID 2033 2043 BY SIMILARITY.
 FT DISULFID 2050 2065 BY SIMILARITY.
 FT DISULFID 2052 2068 BY SIMILARITY.
 FT DISULFID 2070 2080 BY SIMILARITY.
 FT MOD RES 1954 1954 HYDROXYLATION (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 838 838 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1173 1173 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1213 1213 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1308 1308 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1318 1318 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1640 1640 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1704 1704 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1761 1761 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2044 2044 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2173 2173 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2192 2192 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2382 2382 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2472 2472 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2504 2504 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3313 AA; 359348 MW; B1IDA09517288764 CRC64;
 Query Match 11.5%; Score 441; DB 1; Length 3313;
 Best Local Similarity 23.9%; Pred. No. 6.8e-20;
 Matches 187; Conservative 87; Mismatches 256; Indels 252; Gaps 42;
 QY 61 RHPETLALGHE--HTVTLRSITQSSLI-----VGD-----LAPVNGTSQK- 100
 DB 1277 RFUSPL-LGHFLEGVAAVLTEDVFNIQNDTVGGTVLVNVSFALAP-RGAGAA 1334
 QY 101 ---FOGLDNEELYLGYPDYGAIPKAGLSS-----GFTGCVREL 137
 DB 1335 GPFWSSEELQEQLY-----VRAALAARSLDLVLPDDNVCLREPCENYMKCVSVL 1385
 QY 138 R-----LQGEIVFHDNLNATAGISHCP-----TCRDRPCONGQCHDSE 177
 DB 1386 RFDSSAPFLASASTLFRPQPTAGLRCCPFGTODPCETELDLCYNPCNGACARRE 1445
 QY 178 SSSYVCVCPAGFTGSRCE--HSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRS--GLR 234
 DB 1446 -GGYTCVCRPRFTGEDCELDTEAGRCVPGVCVRNGGTCNAEPG-GFRCCQCPAGAFEGPR 1503
 QY 235 CSEGVTVTPSLSCAGSYLALPANTYHRELVDVERKPLADGVLLFSGKSGPVEDFV 294
 DB 1504 CEVAARSFFPS-----SFWNFGRLQRFH-LTSLSLSPATVQPSG-LLFYNGRLNKEHDFL 1556

QY 295 SLAMVGGHLEFRYELGSGLAIVLSAEP--LALGRWHRVSAERLNKDGSLRVNGRPLVRS 352
 DB 1557 ALELVAGVQRLTYSTIGESSTVSTVPGGLSDGQWHTVHLRYNKPRTDALGAQ----- 1611
 QY 353 SPGK-----SGLNLHTLLYLGVEPSV 375
 DB 1612 GFSKDKVAVLSDVDCNVAVALRFAEIGNYSCAAAGVQTSKKSLLDTGFLLLGGV-PNL 1670
 QY 376 PLS-PATNNSAHFRGCVGEVSVNGKRLDLYSFLGSGQIGQCYDSSPCERQPCQHGATCM 434
 DB 1671 PENPFVSRKD--FIGWEDLHIDGRVDM-AFVANNGT-----TAGCQ 1711
 QY 435 PAGRYEFOCLCRDGFKBLCEHENPCQLRBPCLHGTCQ-----GTFCLCLPQFSGRPCQ 490
 DB 1712 AKSHF-----CASG-----PCNKGGLCSERMGGFCDCPVFGGKDCR 1749
 QY 491 QGSGHGIAESWHLEGSG-----GNDAPGQYGAIFHDDGFLAPPGHVFSRSLPEVPETI 544
 DB 1750 LTMAL-----PVHFGNGTSLSWDFGNDMP-----VSVPWYL 1780
 QY 545 ELEVTSTASGLLLWQVEVG-----EAGQKDFISLGLQDGLHYFRYQLSGSEARLV 597
 DB 1781 GLSPFRATKGVLM--QVQLGPHSVLLCKLDQGLLSVTLRSAGHAVH-----LLL 1829
 QY 598 SEDPINDGEWH--RVTLARE--GRRG-----SIQVDGEELVSGRSPGENVAVNAKGSYIG 649
 DB 1830 DQMTVSDGCRWHDRLRLQEEPPGRRGHHIFMVSLDFTLFQDTWAMGSELEGLKYKHLHV 1889
 QY 650 GAPDVATITGGFSSGTCGVKNVLVLSARPAA--PPQPLDLOHR--AAQAN-TRPC 703
 DB 1890 GPPSSKEEG---PQGLVGCIGQVWTGTFPGSGALPPP-----SHRINVEPGCTVTNPC 1941
 QY 704 PS 705
 DB 1942 AS 1943
 RESULT 8
 NX2A_RAT
 ID NX2A_RAT STANDARD; PRT; 1715 AA.
 AC Q63374; Q63375;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Neurexin 2-alpha precursor (Neurexin II-alpha).
 GN NRXN2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., VARIANT LEU-434, AND ALTERNATIVE SPLICING.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RX MEDLINE=92320296; PubMed=1621094;
 RA Ushkaryov Y.A., Petrenko A.G., Geppert M., Suedhof T.C.;
 RT "Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin receptor and laminin.";
 RL Science 257:50-56(1992).
 RN [2]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE=Brain;
 RX MEDLINE=95209856; PubMed=7695896;
 RA Ullrich B., Ushkaryov Y.A., Suedhof T.C.;
 RT "Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons.";
 RL Neuron 14:497-507(1995).
 RN [3]
 RP INTERACTION WITH NEUREXOPHILIN 1.
 RX MEDLINE=99074239; PubMed=9856994;
 RA Missler M., Hammer R.E., Suedhof T.C.;
 RT "Neurexophilin binding to alpha-neurexins. A single LNS domain functions as an independently folding ligand-binding unit.";
 RL

J. Biol. Chem. 273:34716-34723 (1998).
 [4]
 INTERACTION WITH ALPHA-DYSTROGLYCAN.
 MEDLINE=2163578; PubMed=11470830;
 RA Sugita S., Saito F., Tang J., Satz J., Campbell K., Suedhof T.C.;
 "A stoichiometric complex of neuexins and dystroglycan in brain.";
 J. Cell Biol. 154:435-445 (2001).
 CC -1- FUNCTION: Neuronal cell surface protein that may be involved in
 cell recognition and cell adhesion. May mediate intracellular
 signaling.
 CC SUBUNIT: The laminin G-like domain 1 binds to neuexophilin 1.
 CC Isoforms alpha 2C bind to alpha-dystroglycan.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative promoter;
 CC Comment=A number of isoforms, alpha-type (shown here) and
 beta-type, are produced by use of alternative promoters.
 CC Beta-type isoforms differ from alpha-type isoforms in their
 N-terminus;
 CC Event=Alternative splicing; Named isoforms=1;
 CC Comment=At least 216 isoforms may be produced by alternative
 splicing. There is a combination of five alternatively spliced
 extracellular domains at sites 1 to 5, each consisting of
 modular sequences (A-C) that seem to be used independently.
 CC Additional isoforms may derive from a minor cytoplasmic splice
 site 6. Beta isoforms (AC Q63376) share the combination of
 alternatively spliced domains at sites 4, 5 and 6;
 CC Name=Alpha 1A2A3A4A5A;
 CC IsoId=Q63374-1; Sequence=Displayed;
 CC -1- TISSUE SPECIFICITY: Brain (neuronal synapse).
 CC -1- SIMILARITY: Contains 6 laminin G-like domains.
 CC -1- SIMILARITY: Contains 3 EGF-like domains.
 CC -1- SIMILARITY: Belongs to the neuexin family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMIL; M96376; AAA1706.1; -;
 DR EMIL; M96376; AAA1707.1; -;
 DR PIR; C40228; C40228.
 DR HSSP; Q63373; 1C4R.
 DR InterPro; IPR008985; Cona like lec_gl.
 DR InterPro; IPR006209; EGF Like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001791; Laminin G.
 DR InterPro; IPR003585; Neuexin-like.
 DR Pfam; PF00008; EGF; 3.
 DR Pfam; PF00054; laminin G; 5.
 DR SMART; SM00294; 4.1m; 1.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00282; LamG; 6.
 DR PROSITE; PS00022; EGF_1; FALSE NEG.
 DR PROSITE; PS01186; EGF_2; FALSE NEG.
 DR PROSITE; PS00026; EGF_3; 3.
 DR PROSITE; PS50025; LAM G DOMAIN; 6.
 DR Signal; Transmembrane; Repeat; Cell adhesion; EGF-like domain;
 KW Glycoprotein; Alternative splicing; Alternative promoter usage;
 KW Polypeptide.
 FT SIGNAL 1 29 BY SIMILARITY.
 FT CHAIN 30 1715 NEUREXIN 2-ALPHA.
 FT DOMAIN 30 1639 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1640 1660 POTENTIAL.
 FT DOMAIN 1661 1715 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 30 206 LAMININ G-LIKE 1.
 FT DOMAIN 202 242 EGF-LIKE 1.
 FT DOMAIN 289 486 LAMININ G-LIKE 2.
 FT DOMAIN 493 686 LAMININ G-LIKE 3.
 FT DOMAIN 690 727 EGF-LIKE 2.

FT	DOMAIN	732	907	LAMININ G-LIKE 4.
FT	DOMAIN	921	1096	LAMININ G-LIKE 5.
FT	DOMAIN	1099	1136	EGF-LIKE 3.
FT	DOMAIN	1140	1348	LAMININ G-LIKE 6.
FT	DOMAIN	1371	1374	POLY-THR.
FT	DOMAIN	1377	1380	POLY-THR.
FT	DOMAIN	1447	1450	POLY-PRO.
FT	DOMAIN	1647	1650	POLY-ALA.
FT	CARBOHYD	60	60	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	295	295	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	338	338	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	844	844	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1239	1239	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	250	259	Missing (in some isoforms Alpha 1B).
FT	VARSPLIC	250	259	/FtId=VSP_003509.
FT	VARSPLIC	250	283	Missing (in some isoforms Alpha 1C).
FT	VARSPLIC	385	399	/FtId=VSP_003510.
FT	VARSPLIC	393	399	Missing (in some isoforms Alpha 2C).
FT	VARSPLIC	797	806	Missing (in some isoforms Alpha 2B).
FT	VARSPLIC	807	809	Missing (in some isoforms Alpha 3B).
FT	VARSPLIC	1256	1285	Missing (in some isoforms Alpha 4B).
FT	VARSPLIC	1421	1614	Missing (in some isoforms Alpha 5B).
FT	VARSPLIC	1667	1715	Missing (in some isoforms Alpha 6).
FT	VARIANT	434	434	DEGSVQVQDSRNYSNSAQSNGAVVKEKAPAPKTPSKAKK
FT	SEQUENCE	1715	AA; 185282	NKDEYV -> CRKSPREKLLPSGAOLGLDLAKACVC
FT	SEQUENCE	1715	AA; 185282	RCRATCIAGKLEERGSGRGERMQIYIKNK (in
FT	SEQUENCE	1715	AA; 185282	some isoforms Alpha 6).
FT	SEQUENCE	1715	AA; 185282	/FtId=VSP_003517.
FT	SEQUENCE	1715	AA; 185282	N -> L.
FT	SEQUENCE	1715	AA; 185282	59PBF18661F3DB15 CRC64;
FT	SEQUENCE	1715	AA; 185282	11.0%; Score 421; DB 1; Length 1715;
FT	SEQUENCE	1715	AA; 185282	Query Match
FT	SEQUENCE	1715	AA; 185282	Best Local Similarity
FT	SEQUENCE	1715	AA; 185282	21.6%; Pred. No. 5.9e-19;
FT	SEQUENCE	1715	AA; 185282	Matches 193; Conservative 114; Mismatches 274; Indels 312; Gaps 37;
QY	2	IKITERPDSADGMLLYN-GQKRVPGSPNLANROPDFISFGLVGRGPRFRFDAGSGMATI	60	
DB	517	ISLDFRTTEPENGLLFSGQRAGAGVGVSHSSQRADYFAMELLDGLYLLDLMGSGGKIL	576	
QY	61	RHPT-PLALGHFHTVTLRLSLTQGLIIVGDIA-PVNGTSQKFGQGLDNEELYLGSPDY	118	
DB	577	RASSRKVNDGEWCHVDQFQDGRKSGISVNSRSTPLAT--GESEVIDLESELYLGLPEG	634	
QY	119	GAIP-----KAGLSGFIGVRELRIQGBEIVFHDNLNLTAGHSHCP-----TC	162	
DB	635	GRVDLPLPEVWTAALRAGVGVGVRLDFIDGRSRLGLAEAGAGVAVPFCRSRETLKQC	694	
QY	163	RDRPQONGQCHDSSESSVVCV-PAGFTGSGCE-----HSCA	199	
DB	695	ASAPCRNGGICREG-WNRVFCDCIGTGLGRVCEREATVLSYDGSYMKIMLPNAMHTEA	753	
QY	200	-----LHCHEACGPDATC-----	213	
DB	754	EDVSLRPMQRAYGLMMATTSRESADTLELDGGOMKLVNLDCLRVGCAPSAAKGPE	813	
QY	214	-----VNRP-----	217	
DB	814	TLFAGHKLNDNEWHTLRVVRGKSLQLSVNDVTEGQAGAHTRLEFHNIEGTMTERRF	873	
QY	218	-----DGRGYTCRHLG-----RSLRCBEGVTVTTPSLSGAGS	251	
DB	874	ISVPSNFTGHSLGVFNQPTMDQCKDQDITYCELNARFGLRAIVADPVTFKRS--S	930	
QY	252	YLALPALTNTHRELRLDVEFKPLADPVLVLFSGGSGPVEDFVSLAMWGHHLEFRVELGS	311	
DB	931	YLALATL-QAYASMLFFQFKTAPDGLLLFNSGNG---NDFIVIELVRGYHYHVEDLGN	986	

```

312 GLAVLR--SAEPALGRWHRVSAER-LNKDGLSLRVNGRPVLRSPKSGOGLNHLTYL 368
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
987 SPSLMKGNKDPVNDQWNVVSRDPGNVTLKID-SRTVTHSNG-ARNLDLKGELYI 1044
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
369 GGVPSVPLSPATNMAH---FRGCGVEVSVNGKRLDLYSLGSGIGQCY-----DSS 420
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1045 GGLSKNM-FENPKLVASRDGFGGCLASVDLNGRPLDLIADAL--HRIGQVVERGCDGPST 1101
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
421 PCRQPCQHGATCNPAGEVEFYECCLRDGPKGDLCEHSENPOLRPFCLHGFGCGTRCLC 480
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1102 TCTEESCANQVCLQW---DGF-----TCDCT-----1126
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
481 LPQFSGPRCOQ-----GSGHGIAESDWHLESGGNDAPQGVGFHDDGFLAFEGHVF 533
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1127 MTSYGVPVNDPCTTYIFGKGGALITYH-----PNDPSPTR---MDRLAVGESHTQR 1177
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
534 SRSLPEVFTIEVETSTASGLLLMQGVVEGAGQKDFISLQDGHVFRVYOLSGGE 593
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1178 SAVL-----VRVDSASGL-----GDYQLQHDQGTGVGVFNVGTTDD 1213
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
594 ARLVSEDP---INDGEWHRVTLREGRRGSIQVD---GEELVSGRSPQPNVAV-----640
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1214 --ITIDEPNAIVSDGXHVVFRTSGSGNATLQVDSFVNERYPAGNFENRLATARQIP 1271
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
641 -----NAKGSVYIGGAPDVATLTGGRFSSGITGCVN 672
1272 YRLGRVVDEWLDKGRQLTIFNSQAIAIKTGGRDQ-----GRPFQGVSGLYN 1319
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 9

```

NX2A HUMAN
ID NX2A HUMAN STANDARD; PRT; 1712 AA.
AC QP2S2; QY2D6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neurexin 2-alpha precursor (Neurexin II-alpha).
GN NRXN2 OR KIAA0921.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Fetal brain;
RA Seki N., Yoshikawa T., Azuma T., Muramatsu M., Saito T.;
RT "Human neurexin II.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Rowen L., Madan A., Qin S., Baradarani L., Birditt B., Bloom S.,
RA Burke J., Dors M., Fleetwood P., Kaur A., Madan A., Nesbitt R.,
RA Pate D., Hood L.;
RT "Sequencing of human neurexin II gene.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND REVISIONS.
RC TISSUE=Brain;
RX MEDLINE=22158633; PubMed=12168954;
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT "Construction of expression-ready cDNA clones for KIAA genes: manual
RT curation of 330 KIAA cDNA clones.";
RL DNA Res. 9:99-106(2002).
RN [4]
RP SEQUENCE OF 272-1712 FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=9246063; PubMed=10311032;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";

```

```

RL DNA Res. 6:63-70(1999).
CC -1- FUNCTION: Neuronal cell surface protein that may be involved in
CC cell recognition and cell adhesion. May mediate intracellular
CC signaling.
CC -1- SUBUNIT: The laminin G-like domain 1 binds to neurexophilin 1.
CC Specific isoforms bind to alpha-dystroglycan (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative promoter;
CC Comment=A number of isoforms, alpha-type (shown here) and
CC beta-type (AC P58401), are produced by use of alternative
CC promoters. Beta-type isoforms differ from alpha-type isoforms
CC in their N-terminus;
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=Q9P2S2-1; Sequence=Displayed;
CC Name=2; Synonyms=Alpha-2B;
CC IsoId=Q9P2S2-2; Sequence=VSP_003505, VSP_003506, VSP_003507,
CC VSP_003508;
CC -1- TISSUE SPECIFICITY: Predominantly expressed in brain.
CC -1- SIMILARITY: Contains 6 laminin G-like domains.
CC -1- SIMILARITY: Contains 3 EGF-like domains.
CC -1- SIMILARITY: Belongs to the neurexin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB035266; BAA94075.1; -.
CC EMBL; AC044790; AAK68154.1; -.
CC EMBL; AB023138; BAA76765.2; ALT_INIT.
CC HSP; Q63373; 1C4R.
CC Genew; HGNC:8009; NRXN2.
CC MIM; 600566; -.
CC InterPro; IPR008985; Cona_like_lec_gl.
CC InterPro; IPR006209; EGF_like.
CC InterPro; IPR006210; IEGF.
CC InterPro; IPR001791; Laminin G.
CC InterPro; IPR003585; Neurexin-like.
CC Pfam; PF00008; EGF; 3.
CC Pfam; PF00054; laminin G; 5.
CC SMART; SM00294; 4.lm; 1.
CC SMART; SM00181; EGF; 3.
CC SMART; SM00282; LamG; 6.
CC PROSITE; PS00022; EGF_1; FALSE NEG.
CC PROSITE; PS01186; EGF_2; FALSE NEG.
CC PROSITE; PS50026; EGF_3; 3.
CC PROSITE; PS50025; LAM_G_DOMAIN; 6.
CC KW Signal; Transmembrane; Repeat; Cell adhesion; EGF-like domain;
CC Glycoprotein; Alternative splicing; Alternative promoter usage.
CC SIGNAL 1 28
CC CHAIN 29 1712
CC NEUREXIN 2-ALPHA.
CC DOMAIN 29 1636
CC EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 1637 1657
CC POTENTIAL.
CC DOMAIN 1658 1712
CC CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 29 206
CC EGF-LIKE 1.
CC FT DOMAIN 202 242
CC LAMININ G-LIKE 1.
CC FT DOMAIN 289 486
CC LAMININ G-LIKE 2.
CC FT DOMAIN 493 686
CC LAMININ G-LIKE 3.
CC FT DOMAIN 690 727
CC EGF-LIKE 2.
CC FT DOMAIN 732 904
CC LAMININ G-LIKE 4.
CC FT DOMAIN 918 1093
CC LAMININ G-LIKE 5.
CC FT DOMAIN 1096 1133
CC EGF-LIKE 3.
CC FT DOMAIN 1137 1345
CC LAMININ G-LIKE 6.
CC FT DOMAIN 1368 1371
CC POLY-THR.
CC FT DOMAIN 1374 1377
CC POLY-THR.
CC FT DOMAIN 1444 1447
CC POLY-PRO.
CC FT DOMAIN 1644 1647
CC POLY-ALA.

```


FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 338 338 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 841 841 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1236 1236 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 260 283 Missing (in isoform 2).
 FT VARSPLIC 393 399 /FTid=VSP 003505.
 FT VARSPLIC 797 806 /FTid=VSP 003506.
 FT VARSPLIC 1253 1282 /FTid=VSP 003507.
 FT VARSPLIC 1712 AA; 184980 MW; E488EC40ADA1D030 CRC64;
 SQ SEQUENCE 11.0%; Score 420.5; DB 1; Length 1712;
 Query Match
 Best Local Similarity 21.6%; Pred. No. 6.3e-19;
 Matches 192; Conservative 114; Mismatches 275; Indels 309; Gaps 37;
 QY 2 IKITFEDSADGMLLYN-GQKRVPGSPNLANRQPDFISGLVGRPEFRFAGSGMATI 60
 DB 517 ISDFRTTEFNGLLFSQRRAGGAGSHSAGADYFAMELLDGHLYLLLDGSGGKIL 576
 QY 61 RHPT-PLALGHFTVTLRLSLTQSLVGLDIA-PVNGTSGQKFGQGLDNEELVGLGYPY 118
 DB 577 RASSRKVNDGEWCHVDQDRKSGISVNSRSTFFLAT--GDSEILDSELYLGLPEG 634
 QY 119 GAIP-----KAGLSSGPGICVRELRIQGEIVFHDNLNLAHGT-SHCP-----TC 162
 DB 635 GRVDLPPEVETAAALRAGYVGVCRDFTDGRSRDLRGLAAGAGVAVFPSCRETLKQC 594
 QY 163 RDRPCQNGQCHDSSESSYVCV-PAGFTGSRCE-----HSQA 199
 DB 695 ASAPCRNGGVCREG-WNRFCIDCIGTGLGRVCEERATVLSYDGSVMYKIMLPNMAHTEA 753
 QY 200 -----LCHPEACGPD-----210
 DB 754 EDVSLRFMSQRAYGLMMATTRESADTLRLDGGQMKLTNLDCLRVGCAPSKEPTLF 813
 QY 211 -----ATCVNRP-----217
 DB 814 AGHKLNDNEHWITVRVARGSLQLSDVNTVVGQAGAHMRLEFHNITGIMTERRFISV 873
 QY 218 -----DGRYTCRHLG-----RSLGCEGVTVTTPSLSGAGSYLA 254
 DB 874 VPSNFIGLSLVNGQFYMDCQKDGITCYELNARFLRAIVADPVTFKRSR-----SYLA 930
 QY 255 LPALTNTHELRLDVEKPLAPDGVLLFSGKSGPVEDFVSLAMVGGHLEFRVELSGSLA 314
 DB 931 LATL-QAYASHELFPQKTAAPDGLLLFNSGNG---NDFIVIELVKYIHYVFDLNGPS 986
 QY 315 VLR--SAPPLALGRHVRSAER-LNKDGLRVNGRVPVLRSSPKSGQLNLHLLYLGCV 371
 DB 987 LMKGNSDKFVNDQWNNVVRSDPQNVHTLKID-SRTVTQHSNG-ARNLDLKGELYTGL 1044
 QY 372 EPSVPLSPATNMSAH---FRGCVGVSVNGKRLDLYTSLGSGIGQCY-----DSQPC 423
 DB 1045 SKNM-FSNLPLKVASRDGFCGLASVDLNGRLPLDIALD--HRIGQVRCGDPSTTCT 1101
 QY 424 RQPCOHGATCMPAGEYEFQCLCRDGPFGKDLCEHENPQLREPCPLHGGTGTGTRCLCLPG 483
 DB 1102 BESCANQGVCLQW-----DGF-----TCDCT-----MTS 1126
 QY 484 FSGPRCQO-----GSGHGTAESDWHLEGGGNDAPQCYGAYPHDDGFLAFPGHVSRS 536
 DB 1127 YGPGVPCNDPGTYTFGKGGALITYTW-----PPNDRPSTR-----MDRLAVGFSTHQSASV 1177
 QY 537 LPEVETIELEVRTSTAGLLLMQVEVGEAGQKDFISLGLDGHILVFRYQLGSGEARL 596
 DB 1178 L-----VRVDSASGL-----GDYLQLHIDQGTGVIFNVGTDG--I 1211
 QY 597 VSEDP---INDGEHRTVTRALRGGRGSTQVD-----GEELVSGRFGPNNAV-----640
 DB 1212 TIDEPNAIVSDGKHVVVFRFTSGGNATLQVDSWPFVNRYPAGNFNDEKLATARORIPYRL 1271

QY 641 -----NAKGSVYIGGADVDATLTGGRSSGIGTCVKV 672
 DB 1272 GRVDEWLLDKGRLTIFNSQAALKIGRDQ-----GRPFGQVSGLYYN 1316
 RESULT 10
 NX1A_HUMAN
 ID Q9ULB1; O60323; Q9C079; Q9C080; Q9C081; Q9H3M2; Q9UDM6;
 AC Q9ULB1; O60323; Q9C079; Q9C080; Q9C081; Q9H3M2; Q9UDM6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Neurexin 1-alpha precursor (Neurexin I-alpha).
 GN NRXN1 OR KIAA0578.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Seki N., Yoshikawa T., Azuma T., Saito T., Muramatsu M.;
 RT "Human neurexin I-alpha."
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 105-1477 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=98290545; PubMed=9628581;
 RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro."
 RL DNA Res. 5:31-39(1998).
 RN [3]
 RP SEQUENCE OF 441-1121 FROM N.A.
 RA Mead K., Wohldmann P., Minges B.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 1083-1477 FROM N.A.
 RA Rowen L., Madan A., Qin S., Baradarani L., Birditt B., Bloom S.,
 RA Dors M., Dickhoff R., Fleetwood P., Harrison G., Kaur A., Madan A.,
 RA Nesbitt R., Traicoff R., Hood L.;
 RT "Sequencing of human chromosome 2 neurexin I gene."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Neuronal cell surface protein that may be involved in
 CC cell recognition and cell adhesion. May mediate intracellular
 CC signaling.
 CC -!- SUBUNIT: The cytoplasmic C-terminal region binds to CASK. Laminin
 CC G-like domain 2 binds to NKP1. Specific isoforms bind to alpha-
 CC dysglycan and to alpha-latroxin. Binds SYTL1 (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative promoter;
 CC Comment-A number of isoforms are produced by use of alternative
 CC promoters. The alpha and beta (AC P58400) isoforms differ in
 CC their N-terminus;
 CC Event-Alternative splicing; Named isoforms=1;
 CC Comment-A number of isoforms may be produced;
 CC Name=1;
 CC IsoId=Q9ULB1-1; Sequence=Displayed;
 CC -!- TISSUE SPECIFICITY: Heart and brain.
 CC -!- PTM: N- and O-glycosylated (By similarity).
 CC -!- SIMILARITY: Contains 6 laminin G-like domains.
 CC -!- SIMILARITY: Contains 3 EGF-like domains.
 CC -!- SIMILARITY: Belongs to the neurexin family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

alternative splicing and expressed in distinct subsets of neurons.";
 Neuron 14:497-507(1995).
 [2]
 INTERACTION WITH NEUREXOPHILIN.
 MEDLINE=96285455; PubMed=8699246;
 Petrenko A.G., Ullrich B., Missler M., Krasnoperov V., Rosahl T.W.,
 Suedhof T.C.;
 RA "Structure and evolution of neuexophilin.";
 RL J. Neurosci. 16:4360-4369(1996).
 CC -1- FUNCTION: Neuronal cell surface protein that may be involved in
 CC cell recognition and cell adhesion. May mediate intracellular
 CC signaling.
 CC -1- SUBUNIT: The cytoplasmic C-terminal region binds to CASK. Binds
 CC SYTL1. Laminin G-like domain 2 binds to NXP1. Isoforms 9 and
 CC alpha-4c bind to alpha-dystroglycan. Isoform alpha-4c binds to
 CC alpha-latroxin (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -1- ALTERNATIVE PRODUCTS:
 CC -1- Event=Alternative promoter;
 CC Comment=A number of isoforms, alpha-type (shown here) and
 CC beta-type, are produced by use of alternative promoters.
 CC Beta-type isoforms differ from alpha-type isoforms in their
 CC N-terminus;
 CC Event=Alternative splicing; Named isoforms=12;
 CC Comment=Additional isoforms seem to exist. There is probably
 CC more than 96 isoforms. There is a combination of five
 CC alternatively spliced domains at sites 1 to 5, each consisting
 CC of modular sequences (A-G) that seem to be used independently.
 CC For splice site 1 additional splice modules might be possible.
 CC thus increasing the number of possible isoforms. Beta isoforms
 CC (AC Q28142) share the combination of two alternatively spliced
 CC domains at sites 4 and 5. Experimental confirmation may be
 CC lacking for some isoforms;
 CC Name=1; Synonyms=Alpha-1A2A3A4A5A;
 CC IsoId=Q28146-1; Sequence=Displayed;
 CC Name=2; Synonyms=Alpha-1B;
 CC IsoId=Q28146-2; Sequence=VSP_003478;
 CC Name=3; Synonyms=Alpha-1C;
 CC IsoId=Q28146-3; Sequence=VSP_003474, VSP_003478;
 CC Name=4; Synonyms=Alpha-1D;
 CC IsoId=Q28146-4; Sequence=VSP_003477;
 CC Name=5; Synonyms=Alpha-1E;
 CC IsoId=Q28146-5; Sequence=VSP_003474, VSP_003477;
 CC Name=6; Synonyms=Alpha-1F;
 CC IsoId=Q28146-6; Sequence=VSP_003476, VSP_003478;
 CC Name=7; Synonyms=Alpha-1G;
 CC IsoId=Q28146-7; Sequence=VSP_003475;
 CC Name=8; Synonyms=Alpha-2B;
 CC IsoId=Q28146-8; Sequence=VSP_003480;
 CC Name=9; Synonyms=Alpha-2C;
 CC IsoId=Q28146-9; Sequence=VSP_003479;
 CC Name=10; Synonyms=Alpha-3B;
 CC IsoId=Q28146-10; Sequence=VSP_003481;
 CC Name=11; Synonyms=Alpha-4B;
 CC IsoId=Q28146-11; Sequence=VSP_003482;
 CC Name=12; Synonyms=Alpha-5B;
 CC IsoId=Q28146-12; Sequence=VSP_003483;
 CC -1- PTM: N- and O-glycosylated (By similarity).
 CC -1- SIMILARITY: Contains 6 laminin G-like domains.
 CC -1- SIMILARITY: Contains 3 EGF-like domains.
 CC -1- SIMILARITY: Belongs to the neuexin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L14855; AA474123.1; -
 CC PIR; I45944; I45944.
 CC HSSP; Q63373; 1C4R.

DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR008985; ConA_like_1ec_g1.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001791; Laminin G.
 DR InterPro; IPR003585; Neuexin-like.
 DR Pfam; PF00008; EGF_2.
 DR Pfam; PF00054; laminin G; 6.
 DR SMART; SM00294; 4.lm; 1.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00282; LamG; 6.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; FALSE_NEG.
 DR PROSITE; PS00026; EGF_3; 3.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 6.
 KW Signal; Transmembrane; Repeat; Cell adhesion; EGF-like domain;
 KW Glycoprotein; Alternative splicing; Alternative promoter usage.
 FT SIGNAL 1 30
 FT CHAIN 31 1530
 FT DOMAIN 31 1454
 FT TRANSMEM 1455 1475
 FT DOMAIN 1476 1530
 FT DOMAIN 31 212
 FT DOMAIN 213 256
 FT DOMAIN 299 496
 FT DOMAIN 503 695
 FT DOMAIN 699 736
 FT DOMAIN 741 914
 FT DOMAIN 928 1103
 FT DOMAIN 1106 1143
 FT DOMAIN 1149 1347
 FT DOMAIN 1377 1380
 FT DOMAIN 1462 1465
 FT CARBOHYD 125 125
 FT CARBOHYD 190 190
 FT CARBOHYD 813 813
 FT CARBOHYD 1246 1246
 FT VARSPLIC 258 258
 FT VARSPLIC 258 293
 FT VARSPLIC 264 273
 FT VARSPLIC 274 293
 FT VARSPLIC 278 293
 FT VARSPLIC 395 409
 FT VARSPLIC 403 409
 FT VARSPLIC 806 815
 FT VARSPLIC 1263 1292
 FT VARSPLIC 1426 1428
 FT SEQUENCE 1530 AA; 167937 MW; 8A4E4A75C4EC03D1 CRC64;
 SQ
 Query Match 10.6%; Score 405.5; DB 1; Length 1530;
 Best Local Similarity 20.5%; Pred. No. 4.9e-18;
 Matches 203; Conservative 112; Mismatches 285; Indels 391; Gaps 42;
 QY 2 IKITRPDSADGMLLY-NGQKVPVGSPTNLNRPDFTSFLVGVGRPFRRFAGSGMATI 60
 DB 527 ISFDFTTEPNGLLIIFSHGKPRHQKXPMKVDFFAIEMLDGHLVLLDMSGSTIKI 586
 .QY 61 RHPTPLAL-----GHFTVTLLRLSLTQGSLLVGDJ-APVNGTSQKFGGLDNEELVIG 113
 DB 587 K-----ALQKVGNDGEYHVDVFDGRSGTISVNTLRTPY--TAFGESQLDLDDELVIG 639

CC -1- SIMILARITY: Contains 3 EGF-like domains.
CC -1- SIMILARITY: Belongs to the neuexin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to a license@isb-sib.ch).
CC
CC EMBL; M96374; AAA41704.1; --
CC PIR; A40228; A40228.
CC HSP; G63373; 1C4R.
CC InterPro; IPR000152; Aex_hydroxyl_S.
CC InterPro; IPR008985; Con_1-like_1ec_g1.
CC InterPro; IPR006209; EGF_Like.
CC InterPro; IPR006210; IEGF.
CC InterPro; IPR001791; Laminin G.
CC InterPro; IPR003585; Neuexin-like.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF00054; laminin G; 6.
CC SMART; SM00294; 4.lm; 1.
CC SMART; SM00181; EGF; 3.
CC SMART; SM00282; LamG; 5.
CC PROSITE; PS00010; ASX_HYDROXYL; 1.
CC PROSITE; PS00022; EGF_1; FALSE_NEG.
CC PROSITE; PS01186; EGF_2; FALSE_NEG.
CC PROSITE; PS50026; EGF_3; 3.
CC PROSITE; PS50025; LAM_G DOMAIN; 6.
CC Signal; transmembrane; Repeat; Cell adhesion; EGF-like domain;
CC Glycoprotein; Alternative splicing; Alternative promoter usage.
CC
FT CHAIN 31 1514 NEUREXIN 1-ALPHA.
FT DOMAIN 31 1438 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1439 1459 POTENTIAL.
FT DOMAIN 1460 1514 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 31 217 LAMININ G-LIKE 1.
FT DOMAIN 213 255 EGF-LIKE 1.
FT DOMAIN 283 480 LAMININ G-LIKE 2.
FT DOMAIN 487 679 LAMININ G-LIKE 3.
FT DOMAIN 683 720 EGF-LIKE 2.
FT DOMAIN 725 898 LAMININ G-LIKE 4.
FT DOMAIN 912 1087 LAMININ G-LIKE 5.
FT DOMAIN 1090 1127 EGF-LIKE 3.
FT DOMAIN 1133 1331 LAMININ G-LIKE 6.
FT DOMAIN 1361 1364 POLY-THR.
FT DOMAIN 1446 1449 POLY-ALA.
FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 797 797 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1230 1230 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 258 258 Missing (in some isoforms Alpha 1C).
FT VARSPPLIC 258 277 /FtId=VSP_003486.
FT VARSPPLIC 274 277 Missing (in some isoforms Alpha 1G).
FT VARSPPLIC 274 277 /FtId=VSP_003487.
FT VARSPPLIC 379 393 Missing (in some isoforms Alpha 1D).
FT VARSPPLIC 379 393 Missing (in some isoforms Alpha 2C).
FT VARSPPLIC 387 393 /FtId=VSP_003489.
FT VARSPPLIC 387 393 Missing (in some isoforms Alpha 2B).
FT VARSPPLIC 790 799 /FtId=VSP_003490.
FT VARSPPLIC 790 799 DCIRINCNS -> G (in some isoforms Alpha 3B).
FT VARSPPLIC 1247 1276 /FtId=VSP_003491.
FT VARSPPLIC 1410 1411 Missing (in some isoforms Alpha 4B).
FT VARSPPLIC 1410 1411 /FtId=VSP_003492.
FT VARSPPLIC 1410 1411 Missing (in some isoforms Alpha 5B).
FT VARSPPLIC 1410 1411 /FtId=VSP_003493.
SQ SEQUENCE 1514 AA; 166167 MW; D99492302968DE14 CRC64;

Query Match 10.5%; Score 403.5; DB 1; Length 1514;
Best Local Similarity 20.4%; Pred. No. 6.5e-18;

Matches 202; Conservative 113; Mismatches 285; Indels 391; Gaps 42;
QY 2 IKITFRDSDADGMLLY-NGQKRVPGSPTNLANRPDPISFGLVGRDEPRFDAGSGMATI 60
DB 511 ISDFRTTBENGLLFFSGHPRQKDAKHQMKVDFFAIEMLDGLHLYLLLDWGSQTIKI 570
QY 61 RHPFTPLAL-----GHFTVTLLRSLLQGSLLVGDGL-APVNGTSQKFGQGLDNEELYL 113
DB 571 K-----ALQKKVNDGEWYHVDFFDGRSGTISVNTLTPT-TPAGSESEILDLDELLYL 623
QY 114 GYDYGALPKAG-----LSSGFICGVRLEIRIQGBEIVFH-----DNLMTAHSISHC 159
DB 624 GLPE-----NRAGLVFPTEVWVALLNYGVGICIRDLFDGSKDIRQMAETQSTAGVAFSC 679
QY 160 -----PTCRDRPCQGGCHDSESSSYVVCVP-AGFTGSRCE----- 195
DB 680 SRETAKPCLSNPKCKNGMCRDG-WNRVVCDSGTGLGRSCEREATVLSYDGSFMFKIQL 738
QY 196 -----HSQA----- 207
DB 739 FVMVHTEAEVSLRFRSQRAYGILMATTSDSADTLRLDLADGRVKLTVNLDICIRNCNS 798
QY 208 --GPDATCV--NRPDGRGYTCR----- 225
DB 799 SKGPETLPAGYNLNDNEWHVTVVVRGKSLKLTVDQDQAMTGOMAGDHTLRFHNIETGI 858
QY 226 -----CHL-CRSGLR----- 243
DB 859 ITERRYLSSVPSNFIGHLSQTFNGMAYIDCKNGDDIDYCELNARFGRNIADPVFTKT 918
QY 244 PSLGAGSYLALPALTNTHRLDVEFKPLAPDGVLLFSGGKSGPVDFVSLAMVGHL 303
DB 919 KS-----SYVALATL-QAYTSMHLFFQFKTSLDGLILYN--SGDGNDFIWEVLKGYL 969
QY 304 EFRYELGSLAVLR--SAPPLALGRHVRSAER-----LNKDGSLRVNGRPFVLR 352
DB 970 HYVFDLNGANLIKGSNNKPLNDQNHVMTSRDTSNLHTVKIDTKITQITAG----- 1023
QY 353 SPKSGQGLNLTLLYLGCV--EPSVPLSPATMSAHFRCVGVSVNGKRLDLYTFLGS 410
DB 1024 -----ARNLDLSDLYIGVAKETYSKPLVHAKGFGQCLASVDLNGRLPDL----- 1072
QY 411 QGIGQCYDSSPCEROPCOHGATCMPAGEYFQCICRDGFKGDLCEHENPQOLRPPCLHG 470
DB 1073 -----ISDALFCN-----GQIE-----RG--CEGPTTCC--EDSCSNQ 1102
QY 471 GTC-----QGTICLC-LPGFSGPRCQ-----GSGHGIAESDWHLEGSGNDAPQYGA 518
DB 1103 GVCLOQWDGFCDCSMTSFSGLPLNDPQTTVIFSKGGQITKN-----PNDRSTRA- 1156
QY 519 YFHDGFLAFPGHVFSRSLPEVPETIELE---VFTSTASGLLLMQGVGEVGEAGQKDFIS 575
DB 1157 ---DRLAIGF-----STVQKEAVLVRVDSSSL-----GDYLE 1186
QY 576 LGLQDGLVFRYQLGSGEARLVSEDP-INDGEWHRTALREGRGSIQVDG-----EELVS 630
DB 1187 LHIHQKGVKFNKGVNFTDIDAEESNAIINDGKHVVRFTSRGNATLQVDSWPFVRYPA 1246
QY 631 GRSPGNVAV-----NAKGSVYITG----- 650
DB 1247 GNDNERLAIAQRIPYRLGRVVDWELLDKGRQLTIFNSQATIIIGGKEQGFQGLSG 1306
QY 651 -----APDVATLTGRFSSGTTGCVKMLV 674
DB 1307 LYVNGKLVNMAENDANIAIVGNVRLVGEVPSMTTESTATAMOSEMSTSIETITLA 1366
QY 675 LHSARFAPPPQPLDLQHRAQAGANTRPSPS 705
DB 1367 TSTARRGKPTKEPIQSQTDDILVASAECPS 1397

RESULT 13
CLR3_MOUSE

ID AC 091210; Q9ESD0; STANDARD; PRT; 3301 AA.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cadherin EGF LAG seven-pass G-type receptor 3 precursor.
GN CELSR3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGES.
RC STRAIN=C57BL/6;
RX MEDLINE=21839555; PubMed=11850187;
RA Tissir F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;
RT "Developmental expression profiles of Celsr (Flamingo) genes in the
RL mouse."; Mech. Dev. 112:157-160(2002).
RN [2]
RP SEQUENCE OF 2574-3046 FROM N.A., AND DEVELOPMENTAL STAGE.
RX MEDLINE=21534980; PubMed=11677057;
RA Forrester C.J., Little P.F.R.;
RT "The flamingo-related mouse Celsr family (Celsr1-3) genes exhibit
RL distinct patterns of expression during embryonic development."; Mech. Dev. 109:91-94(2001).
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=20253755; PubMed=10790539;
RA Forrester C.J., Barclay J., Rees M., Little P.F.R.;
RT "Chromosomal localization of Celsr2 and Celsr3 in the mouse; Celsr3 is
RL a candidate for the tippy (tip) lethal mutant on chromosome 9."; Mamm. Genome 11:392-394(2000).
CC -!- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Expressed in the CNS and in the eye.
CC -!- DEVELOPMENTAL STAGE: Predominantly expressed in the CNS, the
CC emerging dorsal root ganglia and cranial ganglia. In the CNS,
CC expression is uniform along the rostrocaudal axis. No expression
CC is detected until somite stages. Between E10 and E12, expression
CC is strong in the marginal zone (MZ), and lower in the ventricular
CC zone (VZ). At E15, expression is restricted to the brain and
CC olfactory epithelium. In the brain, it is low in VZ but strong in
CC external fields, particularly those with ongoing migration, such
CC as the telencephalic cortical plate, the olfactory bulb, the
CC cerebellum and the tectum. In the newborn and postnatal stages,
CC expression is high in differentiated neuronal fields.
CC -!- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
CC -!- SIMILARITY: Contains 9 cadherin domains.
CC -!- SIMILARITY: Contains 7 EGF-like domains.
CC -!- SIMILARITY: Contains 2 laminin G-like domains.
CC -!- SIMILARITY: Contains 1 laminin G-like domain.
CC -!- SIMILARITY: Contains 1 GPS domain.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
CC in positions 2575 and 2578.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF427498; AAL25099.1; .
CC EMBL; AF188752; AAG17057.1; ALT_FRAME.
CC MGD; MGI:1858236; Celsr3.
CC InterPro; IPR000152; Asx_hydroxyl_S.
CC InterPro; IPR001216; Cadherin.
CC InterPro; IPR008985; ConA_like_1ec_gl.
CC InterPro; IPR000742; EGF 2.
CC InterPro; IPR006209; EGF_like.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; hormn_receptor.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000203; PKD_cys_rich.
DR Pfam; PF00002; 7tm2; 1.
DR Pfam; PF00002; cadherin; 9.
DR Pfam; PF00008; EGF; 5.
DR Pfam; PF00008; EGF; 5.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF00053; laminin_EGF; 1.
DR Pfam; PF00054; laminin_G; 2.
DR PRINTS; PR00205; CADHERIN.
DR PRINTS; PR00249; GPCRSECRETIN.
DR SMART; SM00112; CA; 9.
DR SMART; SM00181; EGF; 6.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; Hormr; 1.
DR SMART; SM00282; LamG; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00232; CADHERIN_1; 7.
DR PROSITE; PS00268; CADHERIN_2; 8.
DR PROSITE; PS00022; EGF_1; 5.
DR PROSITE; PS01186; EGF_2; 4.
DR PROSITE; PS00026; EGF_3; 6.
DR PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
DR PROSITE; PS00227; G_PROTEIN_RECEP_F2_3; 1.
DR PROSITE; PS00281; G_PROTEIN_RECEP_F2_4; 1.
DR PROSITE; PS00221; GPS; 1.
DR PROSITE; PS00025; LAM_G_DOMAIN; 2.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
KW Developmental protein; Hydroxylation; Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 3301 CADHERIN EGF LAG SEVEN-PASS G-TYPE
FT RECEPTOR 3.
FT DOMAIN 32 2531 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2532 2552 1 (POTENTIAL).
FT DOMAIN 2553 2563 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 2564 2584 2 (POTENTIAL).
FT DOMAIN 2585 2592 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2593 2613 3 (POTENTIAL).
FT DOMAIN 2614 2634 4 (POTENTIAL).
FT TRANSMEM 2635 2655 5 (POTENTIAL).
FT DOMAIN 2656 2673 6 (POTENTIAL).
FT TRANSMEM 2674 2694 7 (POTENTIAL).
FT DOMAIN 2695 2716 8 (POTENTIAL).
FT TRANSMEM 2717 2737 9 (POTENTIAL).
FT DOMAIN 2738 2744 10 (POTENTIAL).
FT TRANSMEM 2745 2765 11 (POTENTIAL).
FT DOMAIN 2766 3301 12 (POTENTIAL).
FT DOMAIN 317 424 CADHERIN 1.
FT DOMAIN 425 536 CADHERIN 2.
FT DOMAIN 537 642 CADHERIN 3.
FT DOMAIN 643 747 CADHERIN 4.
FT DOMAIN 748 849 CADHERIN 5.
FT DOMAIN 850 952 CADHERIN 6.
FT DOMAIN 953 1058 CADHERIN 7.
FT DOMAIN 1059 1160 CADHERIN 8.
FT DOMAIN 1161 1257 CADHERIN 9.
FT DOMAIN 1258 1424 EGF-LIKE 1, CALCIUM-BINDING.
FT DOMAIN 1425 1460 EGF-LIKE 2, CALCIUM-BINDING.
FT DOMAIN 1461 1503 EGF-LIKE 3, CALCIUM-BINDING.
FT DOMAIN 1504 1708 LAMININ G-LIKE 1.
FT DOMAIN 1709 1747 EGF-LIKE 4, CALCIUM-BINDING.
FT DOMAIN 1748 1933 LAMININ G-LIKE 2.
FT DOMAIN 1934 1971 EGF-LIKE 5, CALCIUM-BINDING.
FT DOMAIN 1972 2002 EGF-LIKE 6, CALCIUM-BINDING.
FT DOMAIN 2003 2042 EGF-LIKE 7, CALCIUM-BINDING.
FT DOMAIN 2043 2079 EGF-LIKE 8, CALCIUM-BINDING.

FT DOMAIN 2085 2118 LAMININ EGF-LIKE.
 FT DOMAIN 2468 2520 GPS.
 FT DOMAIN 2720 2724 POLY-LEU.
 FT DISULFID 1370 1381 BY SIMILARITY.
 FT DISULFID 1375 1412 BY SIMILARITY.
 FT DISULFID 1414 1423 BY SIMILARITY.
 FT DISULFID 1430 1441 BY SIMILARITY.
 FT DISULFID 1435 1450 BY SIMILARITY.
 FT DISULFID 1452 1459 BY SIMILARITY.
 FT DISULFID 1468 1479 BY SIMILARITY.
 FT DISULFID 1473 1489 BY SIMILARITY.
 FT DISULFID 1491 1502 BY SIMILARITY.
 FT DISULFID 1715 1726 BY SIMILARITY.
 FT DISULFID 1720 1735 BY SIMILARITY.
 FT DISULFID 1737 1746 BY SIMILARITY.
 FT DISULFID 1939 1950 BY SIMILARITY.
 FT DISULFID 1944 1959 BY SIMILARITY.
 FT DISULFID 1961 1970 BY SIMILARITY.
 FT DISULFID 1974 1985 BY SIMILARITY.
 FT DISULFID 1979 1997 BY SIMILARITY.
 FT DISULFID 1999 2008 BY SIMILARITY.
 FT DISULFID 2016 2029 BY SIMILARITY.
 FT DISULFID 2031 2041 BY SIMILARITY.
 FT DISULFID 2048 2063 BY SIMILARITY.
 FT DISULFID 2050 2066 BY SIMILARITY.
 FT DISULFID 2068 2078 BY SIMILARITY.
 FT MOD RES 1952 1952 HYDROXYLATION (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 838 838 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1173 1173 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1213 1213 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1308 1308 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1318 1318 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1638 1638 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1702 1702 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1759 1759 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2042 2042 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2166 2166 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2185 2185 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2375 2375 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2465 2465 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2497 2497 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 2713 2713 L -> LR (IN REF. 2).
 FT CONFLICT 3024 3024 R -> P (IN REF. 2).
 SQ SEQUENCE 3301 AA; 358455 MW; A6B18F2D7F4DEB6 CRC64;
 Query Match 10.5%; Score 403; DB 1; Length 3301;
 Best Local Similarity 23.0%; Pred. No. 1.7e-17;
 Matches 179; Conservative 91; Mismatches 263; Indels 244; Gaps 40;
 QY 61 RHPTPLALGHF--HTVTLRLSLTQGSII-----LAPVNGTSQK- 100
 DB 1277 RFLSPL-LGHELEGVAALATPTEDVDFIENIQNTDVGTVLVNVSFALAP-RGAGAGAA 1334
 QY 101 ---FQGLDNEELVGGYPDYGAIKAGLAS-----GFIGCVREL 137
 DB 1335 GPFSSSEELQELY-----VRRALARSLLDVLPPDDNVCLREPCENYKCVSVL 1385
 QY 138 R-----IQGEIEVFHDLNLTANGISHCP-----TCDRPCQNGQCHDSE 177
 DB 1386 RFDSSAPFLASTLRFRIQIAGLRCKCPGFTGDFCETELDLCSNPNCRNGGACARRE 1445
 QY 178 SSSYVCVPAGFTSRCE-HSQALHCHPEACGPATCNRPDGGYTCRCHLGRS--GLR 234
 DB 1446 -GGYTCVCRPRFTD--CELDTEAGRCVPGVCRNGGTCINAPNG-GFRCCQCPAGAFEGPR 1501
 QY 235 CEEGVTVTPSLSGAGSYLALPALTNTTHELRDLVEFKPLAPDGVLLFSQGKSGPVDFV 294
 DB 1502 CEVAARSFPDS-----SFVFRGLRQRFH-LTSLSPATVQPSG-LLFYNGRLNEKHDFL 1554
 QY 295 SLAMVGGHLEFRYELGSLAVLRNAP--LALGRHVRSAERLNKQSLRNGRPPVLR 352
 DB 1555 ALELVAGQVRITYSTGESNTVSPVPGGLSDGQWHTVHLRYNKPRTDALGGAQ----- 1609

QY 353 SPGK-----SQGLNLTLLYLGVEPSV 375
 DB 1610 GPSKDKVAVLSVDDCNVAVALQFGAEIGNYSCAAAGVQTSKKSGLDITGTLPLLGCV-PNL 1668
 QY 376 PLS-PATNSAHFRGCVGEVSNKRLDLITYSLFGSQIGQCYDSSPCERQPCQHGATCM 434
 DB 1669 PENFPVSHKD--FIGCMRDHLIDGRMDMA-AFVANNGT-----MAGCQ 1709
 QY 435 PAGYEFQCLCRDGFKGDLCHEENPCQLREPCPLHGTCQ-----GTRCLCLPGSGGRQC 490
 DB 1710 AKSHF-----CASG-----PKXNGFCSERWGFCDCPFVGGKDCR 1747
 QY 491 QSGHGIAESDWHLEGG-----GNDAPGGYGAYFHDGFLAPFGHVSRLSEVPETI 544
 DB 1748 LTMAR-----PYHQNGTSLNDFGND-----MAVSVEPWL 1778
 QY 545 ELEVRTSTASGLLWQGVVEG-----BAGQKDFISLGLQDGLHLYFRYOLGSGEARLY 597
 DB 1779 GLSPRTRATKGLM--QVQLGPHSVLLCKLDRLGILLSVTLNRSAGHTVH-----LLL 1827
 QY 598 SEDPINDGEWH--EVTALRE--GRG-----SIOVDGEELVSGRSPGNVAVNAKGSVVIG 649
 DB 1828 DQMTVDGRWHDLKLELQEEPGRRGHIFMVSLDFTLPQDTMANGGELQGLKVKQLHVG 1887
 QY 650 GAPDVATLTGGRFSGGITGCVKNLVLSARPGAPPPQPLDLQHRQAQAN-TRPCPS 705
 DB 1888 GLPPSSKEEG--HQGLVGCIQGVWIGTFPGSSALLPPSHRVNVPCTVTNFCAS 1941

RESULT 14

SLT2 HUMAN
 ID SLT2 HUMAN STANDARD; PRT; 1529 AA.
 AC O94813; O95710; Q9Y5Q7;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Slit homolog 2 protein precursor (h-Slit-2).
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
 RC TISSUE=Fetal lung;
 RX MEDLINE=99033071; PubMed=9813112;
 RA Itoh A., Miyabayashi T., Ohno M., Sakano S.;
 RT "Cloning and expressions of three mammalian homologues of Drosophila
 slit suggest possible roles for slit in the formation and maintenance
 of the nervous system".
 RL Brain Res. Mol. Brain Res. 52:175-186(1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 3), AND TISSUE SPECIFICITY.
 RC TISSUE=Fetal brain, and Fetal kidney;
 RX MEDLINE=99279238; PubMed=10349621;
 RA Holmes G.P., Negus K., Raman S., Algar E., Yamada T.,
 RA Little M.H.;
 RT "Distinct but overlapping expression patterns of two vertebrate slit
 homologs implies functional roles in CNS development and
 organogenesis".
 RL Mech. Dev. 79:57-72(1998).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2), SEQUENCE OF 1122-1129, FUNCTION,
 RP SUBUNIT AND SUBCELLULAR LOCATION.
 RC TISSUE=Fetal brain;
 RX MEDLINE=99200391; PubMed=10102268;
 RA Brose K., Bland K.S., Wang K.H., Arnott D., Henzel W., Goodman C.S.,
 RA Tessier-Lavigne M., Kidd T.;
 RT "Slit proteins bind Robo receptors and have an evolutionarily
 conserved role in repulsive axon guidance".
 RL Cell 96:795-806(1999).
 CC -!- FUNCTION: Plays a role in axon guidance as a repulsive ligand for

```
CC Robo receptors preventing inappropriate midline crossing.
CC -!- SUBUNIT: Binds robo proteins with high affinity.
CC -!- SUBCELLULAR LOCATION: Secreted. The C-terminal cleavage protein is
CC more diffusible than the larger N-terminal protein that is more
CC tightly cell associated.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=O94813-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O94813-2; Sequence=VSP_050035, VSP_050036;
CC Name=3;
CC IsoId=O94813-3; Sequence=VSP_050036;
CC -!- TISSUE SPECIFICITY: Fetal lung and kidney, and adult spinal cord.
CC Weak expression in adult adrenal gland, thyroid, trachea and other
CC tissues examined.
CC -!- SIMILARITY: Contains 7 EGF-like domains.
CC -!- SIMILARITY: Contains 23 leucine-rich (LRR) repeats.
CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
CC -!- SIMILARITY: Contains 1 laminin G-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB017168; BAA35185.1; -
CC EMBL; AF055585; AAD04309.1; -
CC EMBL; AF133270; AAD25539.1; -
CC HSSP; P00743; ICCF.
CC Genew; HGNC:11086; SLIT2.
CC MIN; 603746; -
CC GO; GO:0005576; C:extracellular; IMP.
CC GO; GO:0045499; F:chemorepellant activity; IMP.
CC GO; GO:0005515; F:protein binding; TAS.
CC GO; GO:0008347; P:glia cell migration; IMP.
CC GO; GO:0007509; P:mesoderm migration; IMP.
CC GO; GO:0008045; P:motor axon guidance; TAS.
CC GO; GO:0007399; P:neurogenesis; TAS.
CC GO; GO:0008038; P:neuronal cell recognition; TAS.
CC GO; GO:0007608; P:olfaction; TAS.
CC InterPro; IPR00152; Asx_hydroxyl_S.
CC InterPro; IPR008985; ConA_like_1ec_g1.
CC InterPro; IPR006207; Cys_knot_C.
CC InterPro; IPR000742; EGF_2.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR006209; EGF_Like.
CC InterPro; IPR003645; FollN.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR001611; LRR.
CC InterPro; IPR000483; LRR_Cterm.
CC InterPro; IPR000372; LRR_Nterm.
CC InterPro; IPR003591; LRR_typ.
CC Pfam; PF00008; EGF; 8.
CC Pfam; PF00054; laminin_G; 1.
CC Pfam; PF00560; LRR; 18.
CC Pfam; PF01463; LRRCT; 4.
CC Pfam; PF01462; LRRNT; 4.
CC SMART; SM00041; CT; 1.
CC SMART; SM00179; EGF_Ca; 2.
CC SMART; SM00274; FOLLN; 2.
CC SMART; SM00282; LamG; 1.
CC SMART; SM00369; LRR_TYP; 8.
CC SMART; SM00082; LRRCT; 4.
CC SMART; SM00013; LRRNT; 4.
CC PROSITE; PS00010; ASX_HYDROXYL; 2.
CC PROSITE; PS01185; CTCK_1; 1.
CC PROSITE; PS01225; CTCK_2; 1.
CC PROSITE; PS00022; EGF_1; 9.
CC PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS50026; EGF_3; 9.
DR PROSITE; PS01187; EGF_CA; 2.
DR PROSITE; PS50025; LAM_G_DOMAIN; 1.
DR Neurogenesis; Glycoprotein; Signal; Alternative splicing;
DR EGF-like domain; Repeat; Leucine-rich repeat.
DR SIGNAL 1 30
DR CHAIN 31 1529 SLIT PROTEIN.
DR CHAIN 31 1121 SLIT PROTEIN N-PRODUCT.
DR CHAIN 1122 1529 SLIT PROTEIN C-PRODUCT.
DR SITE 1121 1122 CLEAVAGE.
DR REPEAT 54 1777 LRR 1.
DR REPEAT 78 101 LRR 2.
DR REPEAT 102 125 LRR 3.
DR REPEAT 126 149 LRR 4.
DR REPEAT 151 173 LRR 5.
DR REPEAT 174 197 LRR 6.
DR REPEAT 199 226 LRR 7.
DR REPEAT 299 322 LRR 8.
DR REPEAT 323 346 LRR 9.
DR REPEAT 347 370 LRR 10.
DR REPEAT 372 394 LRR 11.
DR REPEAT 395 418 LRR 12.
DR REPEAT 511 531 LRR 13.
DR REPEAT 532 556 LRR 14.
DR REPEAT 557 580 LRR 15.
DR REPEAT 582 604 LRR 16.
DR REPEAT 605 628 LRR 17.
DR REPEAT 629 652 LRR 18.
DR REPEAT 753 776 LRR 19.
DR REPEAT 777 799 LRR 20.
DR REPEAT 800 823 LRR 21.
DR REPEAT 824 847 LRR 22.
DR REPEAT 849 872 LRR 23.
DR DOMAIN 918 955 EGF-LIKE 1.
DR DOMAIN 957 996 EGF-LIKE 2.
DR DOMAIN 998 1034 EGF-LIKE 3.
DR DOMAIN 1036 1074 EGF-LIKE 4.
DR DOMAIN 1076 1112 EGF-LIKE 5.
DR DOMAIN 1121 1157 EGF-LIKE 6.
DR DOMAIN 1160 1333 LAMININ G-LIKE.
DR DOMAIN 1332 1368 EGF-LIKE 7.
DR DOMAIN 1453 1528 CTCK
DR DISULFID 922 933 BY SIMILARITY.
DR DISULFID 927 943 BY SIMILARITY.
DR DISULFID 945 954 BY SIMILARITY.
DR DISULFID 961 972 BY SIMILARITY.
DR DISULFID 965 984 BY SIMILARITY.
DR DISULFID 986 995 BY SIMILARITY.
DR DISULFID 1002 1013 BY SIMILARITY.
DR DISULFID 1007 1022 BY SIMILARITY.
DR DISULFID 1024 1033 BY SIMILARITY.
DR DISULFID 1040 1053 BY SIMILARITY.
DR DISULFID 1047 1062 BY SIMILARITY.
DR DISULFID 1064 1073 BY SIMILARITY.
DR DISULFID 1080 1091 BY SIMILARITY.
DR DISULFID 1085 1100 BY SIMILARITY.
DR DISULFID 1102 1111 BY SIMILARITY.
DR DISULFID 1125 1136 BY SIMILARITY.
DR DISULFID 1130 1145 BY SIMILARITY.
DR DISULFID 1147 1156 BY SIMILARITY.
DR DISULFID 1336 1346 BY SIMILARITY.
DR DISULFID 1341 1356 BY SIMILARITY.
DR DISULFID 1358 1367 BY SIMILARITY.
DR CARBOHYD 65 66 N-LINKED (GLCNAC. .) (POTENTIAL).
DR CARBOHYD 185 186 N-LINKED (GLCNAC. .) (POTENTIAL).
DR CARBOHYD 564 564 N-LINKED (GLCNAC. .) (POTENTIAL).
DR CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
DR CARBOHYD 794 794 N-LINKED (GLCNAC. .) (POTENTIAL).
DR CARBOHYD 799 799 N-LINKED (GLCNAC. .) (POTENTIAL).
DR CARBOHYD 1009 1009 N-LINKED (GLCNAC. .) (POTENTIAL).
DR CARBOHYD 1010 1010 N-LINKED (GLCNAC. .) (POTENTIAL).
```


FT CARBOHYD 1019 1019 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1183 1183 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1266 1266 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1300 1300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 258 258 S -> SDEE (in isoform 2).
 FT VARSPLIC 480 487 /FTid=VSP_050035.
 FT VARSPLIC 487 Missing (in isoform 2 and isoform 3).
 FT CONFLICT 226 226 /FTid=VSP_050036.
 FT CONFLICT 607 610 Q -> K (IN REF. 2).
 FT CONFLICT 634 634 SLKT -> KPQN (IN REF. 3).
 FT CONFLICT 634 634 L -> M (IN REF. 3).
 SQ SEQUENCE 1529 AA; 159869 MW; 5D19CC5E7FD461BA CRC64;

Query Match 10.4%; Score 397.5; DB 1; Length 1529;
 Best Local Similarity 29.7%; Pred. No. 1.6e-17;
 Matches 108; Conservative 58; Mismatches 141; Indels 57; Gaps 18;

QY 158 HC-----PTCRDRPCQNGSQCHDSSESSVVCVPAGFTGSRCEHSQAL-----HCHPEAC 207
 Db 1072 HCDIDFDDCQNKCKNGAHCTDA-VNGYTCLCPGYGKCEKLVSV---NFINKESYLTQPSAKVRPQT 1182

QY 208 GPDATC-----VNRPDGRGYTCRHLGRGLRCEEGVTTPSLSGAGSYLALPAL-----T 259
 Db 1183 QNGAQCIVRINEP-----ICQCLPGYQGEKCEKLVSV---NFINKESYLTQPSAKVRPQT 1182

QY 260 NTHHELRLDVEFKPLAPDGVLLPFGSGKSPVEDFVSLAMVGGHLEFPEYELSGSLA-VLRS 318
 Db 1183 NITLQIATDED-----SGILLYKGDK-----DHIAVELYGRVRSYDTGSHPASAIYS 1231

QY 319 AEPLALGRWHRVASBELNKGSLRVNGRPPVLRSPSGKSGQNLNLTLLYLGGVB-----P 373
 Db 1232 VETINDGNFHVIELLALQSLSLSDGNGPKIITNLSKQSTLNFDSPLYVGMGPKGNVA 1291

QY 374 SVPLSPATMGAHFGCGVEVSNKRLDLTLYPSLGSQIGQCVDSPPCERQPCQHGATC 433
 Db 1292 SLRQAPQGN-GTSPFGCIRNLYNSELQDFQKVPMTGILPGC---EPCHKVCAHG-TC 1346

QY 434 MPAGEYEFQCLDRGDKGDLCEHEN-PCQLREPCILHGTGTCQ-----GTRCLCLPFGSGP 487
 Db 1347 QSSAGTCTCEQSGWGMFLCDQRTNDFC-LGNKCVH-GTCLPINAFYSYCKCLEHGGV 1404

QY 488 RCOQ 491
 Db 1405 LCDE 1408

RESULT 15

CLR3_HUMAN STANDARD; PRT; 3312 AA.
 AC Q9NYQ7; O75092;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cadherin EGF LAG seven-pass G-type receptor 3 precursor (Flamingo homolog 1) (hml1) (Multiple epidermal growth factor-like domains 2) (Epidermal growth factor-like 1).
 DE CELSR3 OR CDHF11 OR FM11 OR EGFL1 OR MEGF2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=20202599; PubMed=10716726;
 RA Wu Q., Maniatis T.;
 RT "Large exons encoding multiple ectodomains are a characteristic feature of protocadherin genes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).
 RN [2]
 RP SEQUENCE OF 1954-3312 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=98360089; PubMed=9693030;

RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
 RT "Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening.";
 RL Genomics 51:27-34(1998).
 CC -!- FUNCTION: Receptor that may have an important role in cell/cell signaling during nervous system formation.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
 CC -!- SIMILARITY: Contains 9 cadherin domains.
 CC -!- SIMILARITY: Contains 8 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 laminin G-like domains.
 CC -!- SIMILARITY: Contains 1 laminin EGF-like domain.
 CC -!- SIMILARITY: Contains 1 GPS domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
 CC -----
 DR ENBL; AF231023; RAF61929.1; --
 DR ENBL; AS011536; BAA32464.1; --
 DR HSP; P00740; LEDM.
 DR Genew; HGNC:3230; CELSR3.
 DR MIM; 604264; --
 DR GO; GO:0005198; F:structural molecule activity; NAS.
 DR InterPro; IPR000152; Asx hydroxyl_S.
 DR InterPro; IPR002136; Cadherin.
 DR InterPro; IPR008985; ConA_like_lec_gl.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR000832; GPCR secretin.
 DR InterPro; IPR001879; hoxnm_receptor.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001751; Laminin_G.
 DR InterPro; IPR000203; PKD_cys_rich.
 DR Pfam; PF00002; 7tm_2; 1.
 DR Pfam; PF00028; cadherin; 9.
 DR Pfam; PF00008; EGF; 5.
 DR Pfam; PF01825; GPS; 1.
 DR Pfam; PF02793; HRM; 1.
 DR Pfam; PF00054; laminin_G; 2.
 DR PRINTS; PR00205; CADHERIN.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR00249; GPCRSECRETIN.
 DR SMART; SM00112; CA; 9.
 DR SMART; SM00181; EGF; 6.
 DR SMART; SM00303; GPS; 1.
 DR SMART; SM00008; Hoxmr; 1.
 DR SMART; SM00282; Lamg; 2.
 DR PROSITE; PS00010; ASX HYDROXYL; 1.
 DR PROSITE; PS00232; CADHERIN 1; 7.
 DR PROSITE; PS00268; CADHERIN_2; 8.
 DR PROSITE; PS00022; EGF 1; 6-
 DR PROSITE; PS01186; EGF_2; 4.
 DR PROSITE; PS00026; EGF_3; 6.
 DR PROSITE; PS00649; G PROTEIN RECEPT_F2_1; FALSE NEG.
 DR PROSITE; PS00650; G PROTEIN RECEPT_F2_2; FALSE NEG.
 DR PROSITE; PS0027; G PROTEIN RECEPT_F2_3; 1.
 DR PROSITE; PS0261; G PROTEIN RECEPT_F2_4; 1.
 DR PROSITE; PS0221; GPS; 1.
 DR PROSITE; PS00025; LAM G DOMAIN; 2.
 DR PROSITE; PS00025; LAMININ TYPE_EGF; 1.
 DR G-protein coupled receptor; transmembrane; Glycoprotein;
 KW EGF-like domain; Calcium-binding; laminin EGF-like domain; Repeat;
 KW Developmental protein; Hydroxylation; Signal.
 FT SIGNAL 1 32 POTENTIAL.
 FT CHAIN 33 3312 CADHERIN EGF LAG SEVEN-PASS G-TYPE RECEPTOR 3.

FT DOMAIN 33 2540 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2541 2561 1 (POTENTIAL).
FT DOMAIN 2562 2571 CYTOLASMIC (POTENTIAL).
FT TRANSMEM 2573 2593 2 (POTENTIAL).
FT DOMAIN 2594 2601 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2602 2622 3 (POTENTIAL).
FT DOMAIN 2623 2643 CYTOLASMIC (POTENTIAL).
FT TRANSMEM 2644 2664 4 (POTENTIAL).
FT DOMAIN 2665 2681 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2682 2702 5 (POTENTIAL).
FT DOMAIN 2703 2725 CYTOLASMIC (POTENTIAL).
FT TRANSMEM 2726 2746 6 (POTENTIAL).
FT DOMAIN 2747 2753 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2754 2774 7 (POTENTIAL).
FT DOMAIN 2775 3312 CYTOLASMIC (POTENTIAL).
FT DOMAIN 336 433 CADHERIN 1.
FT DOMAIN 434 545 CADHERIN 2.
FT DOMAIN 546 651 CADHERIN 3.
FT DOMAIN 652 756 CADHERIN 4.
FT DOMAIN 757 858 CADHERIN 5.
FT DOMAIN 859 961 CADHERIN 6.
FT DOMAIN 962 1067 CADHERIN 7.
FT DOMAIN 1068 1169 CADHERIN 8.
FT DOMAIN 1170 1265 CADHERIN 9.
FT DOMAIN 1266 1433 EGF-LIKE 1, CALCIUM-BINDING.
FT DOMAIN 1434 1471 EGF-LIKE 2, CALCIUM-BINDING.
FT DOMAIN 1472 1514 EGF-LIKE 3, CALCIUM-BINDING.
FT DOMAIN 1515 1719 LAMININ G-LIKE 1.
FT DOMAIN 1720 1758 EGF-LIKE 4, CALCIUM-BINDING.
FT DOMAIN 1759 1944 LAMININ G-LIKE 2.
FT DOMAIN 1945 1982 EGF-LIKE 5, CALCIUM-BINDING.
FT DOMAIN 1983 2020 EGF-LIKE 6, CALCIUM-BINDING.
FT DOMAIN 2021 2053 EGF-LIKE 7, CALCIUM-BINDING.
FT DOMAIN 2054 2090 EGF-LIKE 8, CALCIUM-BINDING.
FT DOMAIN 2091 2131 LAMININ EGF-LIKE.
FT DOMAIN 2132 2159 GPS.
FT DOMAIN 2160 2177 BY SIMILARITY.
FT DISULFID 1379 1390 BY SIMILARITY.
FT DISULFID 1391 1421 BY SIMILARITY.
FT DISULFID 1422 1432 BY SIMILARITY.
FT DISULFID 1433 1450 BY SIMILARITY.
FT DISULFID 1451 1459 BY SIMILARITY.
FT DISULFID 1460 1470 BY SIMILARITY.
FT DISULFID 1471 1490 BY SIMILARITY.
FT DISULFID 1491 1500 BY SIMILARITY.
FT DISULFID 1501 1513 BY SIMILARITY.
FT DISULFID 1514 1737 BY SIMILARITY.
FT DISULFID 1738 1746 BY SIMILARITY.
FT DISULFID 1747 1757 BY SIMILARITY.
FT DISULFID 1758 1961 BY SIMILARITY.
FT DISULFID 1962 1970 BY SIMILARITY.
FT DISULFID 1971 1981 BY SIMILARITY.
FT DISULFID 1982 1996 BY SIMILARITY.
FT DISULFID 1997 2008 BY SIMILARITY.
FT DISULFID 2009 2019 BY SIMILARITY.
FT DISULFID 2020 2040 BY SIMILARITY.
FT DISULFID 2041 2052 BY SIMILARITY.
FT DISULFID 2053 2074 BY SIMILARITY.
FT DISULFID 2075 2099 BY SIMILARITY.
FT DISULFID 2100 2107 BY SIMILARITY.
FT DISULFID 2108 2117 HYDROXYLATION (POTENTIAL).
FT MOD RES 1963 1963 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 632 632 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 847 847 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1182 1182 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1222 1222 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1317 1317 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1327 1327 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1649 1649 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1713 1713 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1770 1770 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2053 2053 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2177 2177 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2196 2196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2386 2386 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 2474 2474 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2506 2506 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 2158 2158 G -> GLRGAG (IN REF. 2).
SQ SEQUENCE 3312 AA; BEC208703651A4A5 CRC64;
Query Match 10.4%; Score 397.5; DB 1; Length 3312;
Best Local Similarity 23.8%; Pred. No. 3.8e-17;
Matches 174; Conservative 83; Mismatches 252; Indels 221; Gaps 39;
QY 90 LAPVNGTSQK-----FQGLDINEELVILGGYPDYGAIPKAGLSS-----VRRALAAARSLLDVLPFDNDVCLURE 128
DB 1333 LAP-RGAGAGAPMFSSBELOEQLY-----VRRALAAARSLLDVLPFDNDVCLURE 1382
QY 129 ---GFIGCVREL-----IQGEIIVFHDNLTAHGHSHCP-----TCRDR 165
DB 1383 PCENYKCVSVLRFDSAPFLASASTLPRIQPIAGLRCPFGTDFCETELDLCYEN 1442
QY 166 PCONGQCHDSSESSIVCVCPAGFTGSRCE-HSQALHCHPEACGPDATCNVPCDGRGYTC 224
DB 1443 PCRNGGACARRE-GGYTCVCRPRFTGCEDELTEAGRCVGVCRNGGTCTDPNG-GFRC 1500
QY 225 RCHLGRS--GLRCEGVTVTTPSLSGAGSYLALPALTNTHHEIRLDVEFKPLADQVLLF 282
DB 1501 QCPAGAGFEGPRCEVAARSPPS-----SFVNFGRQRQFH-LTSLSEATVQSQG-LLF 1553
QY 283 SGKSGGVEDFVSLAMVGGHLEPRYELGSLAVLRSARP--LALGRHVSARLNK--- 337
DB 1554 YNRLNEKHDFLALVAGQVRLTYSTGESNTVSTVFGGLSDGQMHVTHLYYKRPRT 1613
QY 338 -----DGSRLVNGRPRV-----LRSSPGKSOGLNHLTL 366
DB 1614 DALGGAQPSKDVAVLSDVDDCDVALQFAGIENYSCAAAGVQTSKKS--LDLTGL 1671
QY 367 YLGSVEPSVPLS-PATNMSAHFRGCVGEVSNKRLDLYFLGSGOIGOCYDSSPCERQ 425
DB 1672 LLGGV-PNLPEPFVSHKD--FIGCKEDLHIDGRVDM-AFVANNG-----TMA 1717
QY 426 PCQHGATCMPAGEVEFQCLCRDGFKGLCEHENPCQLREPLHGGTCQ-----GTRCLCL 481
DB 1718 GCQ-----AKLHF--CDSG-----PKNSGFCSERWGFSFDCDP 1749
QY 482 PGFSGPRCQSGSGHGIABSDHLEGGG-----GNAPQYQYGFHDDGLAFPGHVSFR 535
DB 1750 VFGGKDCQLTMAH-----PHFRNGNTLSWFGSD----- 1780
QY 536 SLPEVPETIELEVRTSTAGLLMQGVEGVEAGQKDFISLGLQDGHVFRYQLGSGEAR 595
DB 1781 MAVSPWYLGAFRATQGVIM--QVQAGP-----HSTILCOLDRGLLSVTVTGSGRAS 1834
QY 596 --LVSEDPINDGEWH--RVTLARE--CRRG-----SIOVDGEELVSGRSPGNVAVNAKGS 645
DB 1835 HLLLDQVTVSDGRWHDLRLLEQEPGGRGHVLMVSLDFSLFQDTMAVGSSELQGLKVKQ 1894
QY 646 VYIGAGPDVATLTGFRSSGTTGCVKMLVLSARPAPGAP---PPOPLDLQHRAQA----- 696
DB 1895 LHVGGLPEP---SABEAPQGLVGCIGVWLGSTPSPGFALLPP-----SHVNAEPGCW 1946
QY 697 --GANTRPCEP 704
DB 1947 TNACASGFCP 1956

Search completed: March 9, 2004, 17:19:43
Job time : 22.0517 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:15:08 ; Search time 61.7389 Seconds
(without alignments)
3602.917 Million cell updates/sec

Title: US-10-006-011a-3

Perfect score: 3825

Sequence: 1 EIKITRPDSADGMLLYNQ.....QLDLQRAQAGANTRPCPS 705

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mnc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	901.5	23.6	2026	4	O00468	O00468 homo sapien
2	840	22.0	775	4	Q8N7Y0	Q8N7Y0 homo sapien
3	820	21.4	1009	11	Q8BGP3	Q8BGP3 mus musculu
4	812.5	21.2	708	5	Q9NFS9	Q9NFS9 drosophila
5	812.5	21.2	1035	5	Q9NEG1	Q9NEG1 drosophila
6	812.5	21.2	4117	5	Q8IRV9	Q8IRV9 drosophila
7	812.5	21.2	4179	5	Q9W4Y4	Q9W4Y4 drosophila
8	812.5	21.2	4223	5	Q8WPN3	Q8WPN3 drosophila
9	812.5	21.2	4228	5	Q8IRV8	Q8IRV8 drosophila
10	703	18.4	488	4	Q961C1	Q961C1 homo sapien
11	664.5	17.4	1361	5	Q9V714	Q9V714 drosophila
12	664.5	17.4	1361	5	Q9NGV2	Q9NGV2 drosophila
13	628	16.4	463	4	Q8NAL2	Q8NAL2 homo sapien
14	581.5	15.2	822	11	Q8BU56	Q8BU56 mus musculu
15	576.5	15.1	1039	5	Q9VQB1	Q9VQB1 drosophila
16	511	13.4	925	5	Q9U7E8	Q9U7E8 caenorhabdi

17	479	12.5	350	11	Q8K326	Q8K326 mus musculu
18	464.5	12.1	478	5	Q86SD6	Q86SD6 ciona intes
19	420.5	11.0	1458	11	Q8CJG8	Q8CJG8 rattus norv
20	419.5	11.0	1531	11	Q88279	Q88279 rattus norv
21	412.5	10.8	1531	11	Q9WVB5	Q9WVB5 mus musculu
22	412.5	10.8	1557	11	Q80TR4	Q80TR4 mus musculu
23	411.5	10.8	1530	11	Q9WUG5	Q9WUG5 rattus norv
24	407.5	10.7	1474	11	Q8CJG9	Q8CJG9 rattus norv
25	407	10.6	1534	4	Q75093	Q75093 homo sapien
26	407	10.6	1518	4	Q9U1L7	Q9U1L7 homo sapien
27	402.5	10.5	1576	11	Q80Y87	Q80Y87 mus musculu
28	402.5	10.5	1553	11	Q8CHB6	Q8CHB6 mus musculu
29	394.5	10.3	1523	4	Q75094	Q75094 homo sapien
30	393.5	10.3	517	5	Q8MRJ7	Q8MRJ7 drosophila
31	390	10.2	1512	13	Q9DE36	Q9DE36 brachydanio
32	387.5	10.1	1025	11	Q9Z166	Q9Z166 mus musculu
33	387.5	10.1	1521	11	Q9L1B9	Q9L1B9 mus musculu
34	387	10.1	1523	11	Q88280	Q88280 rattus norv
35	385.5	10.1	950	13	Q90Z44	Q90Z44 gallus gall
36	385.5	10.1	1392	4	Q8IUE3	Q8IUE3 homo sapien
37	385	10.1	1523	11	Q9WVB4	Q9WVB4 mus musculu
38	380.5	9.9	313	13	Q9YHFO	Q9YHFO rana pipien
39	377.5	9.9	1529	13	Q7ZX12	Q7ZX12 xenopus lae
40	375	9.8	738	13	Q90Z45	Q90Z45 gallus gall
41	370.5	9.7	1100	11	Q8CCT8	Q8CCT8 mus musculu
42	367	9.6	1515	13	Q9DE37	Q9DE37 brachydanio
43	363	9.5	1837	5	Q9VCZ9	Q9VCZ9 drosophila
44	362.5	9.5	4307	5	Q19319	Q19319 caenorhabdi
45	361.5	9.5	792	13	Q90Z43	Q90Z43 gallus gall

ALIGNMENTS

RESULT 1
O00468 PRELIMINARY; PRT; 2026 AA.
ID O00468
AC O00468;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE AGRIN precursor (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Groffen A.J.A., Buskens C.A.F., Van Kuppevelt T.H.M.S.M.,
RA Veerkamp J.H., Monnens L.A.H., van den Heuvel L.P.W.J.;
RL Eur. J. Biochem. 0:0-0(1998).
RN [2]
RP SEQUENCE OF 1-153 FROM N.A.
RX MEDLINE=96241170; PubMed=8617505;
RA Lennon G., Auffray C., Polymeropoulos M., Soares M.B.;
RA "The I.M.A.G.E. Consortium: an integrated molecular analysis of
RT genomes and their expression."
RL Genomics 33:151-152(1996).
RN [3]
RP SEQUENCE OF 1-153 FROM N.A.
RA Denzer A.J., Brandenberger R., Gesemann M., Chiquet M., Ruegg M.A.;
RL J. Cell Biol. 0:0-0(0).
RL EMBL; AF016903; AAC39776.1; -
DR EMBL; U84406; AAB52917.1; -
DR HSSP; P00740; 1EDM.
DR Genew; HGNC:329; AGRN.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR004850; Agrin_NtA.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR003645; FcLN.
DR InterPro; IPR002350; Kazal.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.

DR InterPro; IPR000082; SEA domain.
DR InterPro; IPR008993; TIMP like.
DR InterPro; IPR001455; UPF0033.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00050; kazal; 9.
DR Pfam; PF00053; laminin_EGF; 2.
DR Pfam; PF00054; laminin_G; 3.
DR Pfam; PF03146; NCA; 1.
DR Pfam; PF03390; SEA; 1.
DR PRINTS; PRO0011; EGF_LAMININ.
DR SMART; SM00180; EGF_Lam; 2.
DR SMART; SM00274; FOLN; 5.
DR SMART; SM00280; KAZAL; 9.
DR SMART; SM00282; LamG; 3.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; 6.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01248; LAMININ TYPE EGF; 1.
DR PROSITE; PS00025; LAM_G DOMAIN; 3.
DR PROSITE; PS00024; SEA; 1.
DR PROSITE; PS01148; UPF0033; 1.
KW EGF-like domain; Laminin EGF-like domain; Signal.
FT NON_TER 1
FT SIGNAL <1 10 POTENTIAL.
FT CHAIN 11 2026 AGRN.
SQ SEQUENCE 2026 AA; 212881 MW; 4AB0EE710CD4B8EF CRC64;

Query Match 23.6%; Score 901.5; DB 4; Length 2026;
Best Local Similarity 32.2%; Pred. No. 3.2e-58;
Matches 242; Conservative 93; Mismatches 266; Indels 151; Gaps 22;

QY 2 IKITRPSADGMLYNGQKRVGSPPTNLANQPPFISGLVGGPPEPRFAGSGMATIR 61
DB 1377 LALEFRALPEPQGLLYNGNA-----RGKDFLALALDDGRVQLRFTGSGPAVL 1425

QY 62 HPTPLALGHFTVTLRLSTQGLVGLDLPVNGTSQKFGQLDLNEELYLGGYPD--Y 118
DB 1426 SAVPEFGQWHLRSLRHRWRTLSVDGETPVLGSPSGTGLNDLDTLFGVGYPEDQAA 1485

QY 119 GAIPKAGLSGFTGCVRELIQGEIEIVFDLNL-----TAHGISHC--PTCRDRPCQNG 170
DB 1486 VALERTFVAGLRCIRLLDNNQRL---ELGIPGAATRGSGVGECDHPCLPNPCHGG 1542

QY 171 GQCHDSESSVYVCVCPAGTGRCEHSQALHCHPEACGPDATCNVRPDGRGYTCRCHLGR 230
DB 1543 APCQNLEAGRFHQCCPFGVGTCADEKS--PQPNCHGAACPRVLPEG-GAQCCEPLGR 1600

QY 231 SGLRCEGVTVTPSLSGAGSYLALPALNTNTHELR-----LDVEFKPLAP 276
DB 1601 EGTFCQ-----TASQDGGSPFLA-DFNGFSLHLRGLHTFARDLGEKMALEWFLARGP 1654

QY 277 DGVLLFSGKSGVPEDFVSLAMVGGHLEFRYELGSLAVLRSAEPLALGRHVRVSAERLN 336
DB 1655 SGLLYNGQKTDGKDFVSLARDRLREPRYDLGKGAAVIRSEPVTLGAWTRVSLERN 1714

QY 337 KQGLRVNGRPVLRSPGKSGQNLNLTLILYGGVPSVPLSPATNMSAHFRGCVGEYSV 396
DB 1715 RKALAVGDGPRVLGSPVPHVTNLKEPLVYGGADPFSKLARAAVSSGFDGAIQLVSL 1774

QY 397 NGKRL-----DLTVSFLSGQIGQCYDSSPCR---QPCHGATCPAGSEYEQC 443
DB 1775 GGRQLLTPEHVRQVDVT-SFAG-----HPCTRASGHPCNLGASCVP----- 1815

QY 444 LCRDGFKGLDCEHENPCQLREFCLGGTCCQTRCLCLPFGSGPRCQSGHGAESDWH 503
DB 1816 -----REAY-----VCLCPGFGSPGCKGL----- 1837

QY 504 LEGSGNDAPQGYGVAFHDGDTAFPGHVSRLPEVPET-----LELVRSITAG 555
DB 1838 VEKSAG-----DVTTLAFDGRTEVEYLVNAVTESEKALQSNHFLSLRTEATQG 1885

QY 556 LLLWQGVVEVAGQKDFISLGLQDGLVFRYOLGSGEARLVSEDPINDGEHVRVTLRE 615
DB 1838 VEKSAG-----DVTTLAFDGRTEVEYLVNAVTESEKALQSNHFLSLRTEATQG 1885

DB 1886 LVLMWS-----GKATERADYVALAIVDGHQLSLYNLGSQPVLRLSTVPVNTNRLRVVAHRE 1941
QY 616 GRGSIQVDGCELYSGRSPGNVAVNAKSVYIGAPD--VATLTGGRFSSGIGTCVKNL 673
DB 1942 QREGSLQVNGNEAPVTGSSPLGATQDLDGALWIGGPELPVGPALPKAYGTGFGVGLADV 2001
QY 674 VLHSARPAGAPPQPLDLQHRAQAGANTRPCPS 705
DB 2002 VVGR-----HPLHLEDAVTKPELRPCPT 2025

RESULT 2
Q8N7Y0 PRELIMINARY; PRT; 775 AA.
AC Q8N7Y0; 01-OCT-2002 (TriEMBLrel. 22, Created)
DT 01-OCT-2002 (TriEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TriEMBLrel. 25, Last annotation update)
DE Hypothetical protein FL040230.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
RA Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.,
RA "NEDO human cDNA sequencing project".
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK097549; BAC05096.1; -
GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR008985; ConA like lec_gl.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00054; laminin_G; 3.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00179; EGF_Ca; 2.
DR SMART; SM00282; LamG; 3.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS00025; LAM_G DOMAIN; 3.
KW Hypothetical protein; EGF-like domain.
SQ SEQUENCE 775 AA; 84800 MW; 61A81294F0204ED2 CRC64;

Query Match 22.0%; Score 840; DB 4; Length 775;
Best Local Similarity 29.6%; Pred. No. 3.6e-54;
Matches 214; Conservative 107; Mismatches 257; Indels 144; Gaps 20;

QY 1 EKITRPSADGMLYNGQKRVGSPPTNLANQPPFISGLVGGPPEPRFAGSGMATI 60
DB 176 QITLFRFAEADGILLYCGENE-----HGRGDFMSLAIRRSIQFRNCGTGVAIL 226

QY 61 RHPTPLALGHFTVTLRLSTQGLVGLDLPVNGTSQKFGQLDLNEELYLGGYPD-YG 119
DB 227 VSETKILGWHITVLRDGLNGLLQLNNGTPVYQSGQYKSTTPTPLVGGAPSAW 286

QY 120 AIPKAGLSGFTGCVRELIQGEIEIVFDLNL-----LTAHGISHCPT--CRDRPCQNG 170
DB 287 LVRTATGNRGQGVQSLAVNGRRI---DMRPWPLGKALSGADVGECSGICDEASCIHG 343

QY 171 GQCHDSESSVYVCVCPAGTGRCEHSQALHCHPEACGPDATCNVRPDGRGYTCRCHLGR 230
DB 344 GTCTAIAKADSYICLCPLOF----- 362

QY 231 SGLRCEGVTVTPSL-SGAGSYLALPALNTNTHEL---RLDVEFKPLADPGVLLFS--- 283

```
Db 363 KRRHCEAFTIPQFRESLSYAAATPWPLEPQHVLSFMFEIITFRDSDGCVLLIYSDT 422
Qy 284 GKSGPVEDFVSLAMVGHLFRYELGSLAVLSAEPLALGRWHRVSAERLNKDGSLRV 343
Db 423 GSK-----DFLSINLAGHVEFRFCGSGTGLRSEDPLTLGNWHELVRSTAKNGILQV 477
Qy 344 NGRPVLSSPCKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKLDL 403
Db 478 DKQKIVEGMAEGFTQIKCNTDIFIGVNPVDDVKNSGVLPKPSGSIQKIIINDRIHV 537
Qy 404 TYSFLGSGIGQCYDSSPCERQPCQHGATCMPAGEYEQICLRDGFKGDLCHEHENPCQL 463
Db 538 KHDF--TSGVNVENAAHPCVRAPCAHGSCRPKE-GYDCDCPLGFEG-----582
Qy 464 REPCLHGTCQGT--RCILCPGSPRCQCGSGHGIAESDWHLESGGNDAPGQYGFH 521
Db 583 ----LH--CQKAITALEIPOFIG-----RYLYIDN 608
Qy 522 DGFALAFPQHVFRSLPEVPEETIEVLTSTASGILLWQGVGAGQKDFISLGLQDG 581
Db 609 PILKRVSG--SRS-----NVPMRFKTTAKDGLLNRG--DSPMRPNSDFISLGLRDG 657
Qy 582 HLVFRYQLGSGEARLVSDEPINDGEWHRVTALREGRGSIQVDCGELYSRSPGNVAVN 641
Db 658 ALVFSYNLGSGVASIMVNGSFDNGRWHRYKAVRDGSGKITVDYGAETGKSPGMROLN 717
Qy 642 AKGSVYIGAPDVATLTGGRFSGSITGCVKNLVLSRSPGAPPQPLDLQHRAQAGANTR 701
Db 718 INGALYVGGMKEIALHTNRQYMRGLVGCISSHFTLST-----DVHISLVEDAVDGKNIN 770
Qy 702 PC 703
Db 771 TC 772

RESULT 3
Q8BGP3 PRELIMINARY; PRT; 1009 AA.
AC Q8BGP3;
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Forelimb, Skin, Testis, and Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
  60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK031136; BAC27271.1; -
DR EMBL; AK033332; BAC28235.1; -
DR EMBL; AK037223; BAC29762.1; -
DR EMBL; AK041546; BAC30982.1; -
DR MGD; MGI:2146149; AU040377.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR00742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR06209; EGF_like.
DR InterPro; IPR003962; FnIII subd.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001791; Laminin_G.
```

```
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00064; laminin_G; 3.
DR PRINTS; PR00014; ENTPEI11.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00282; LamG; 3.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS50025; Lam_G_DOMAIN; 3.
RW Hypothetical protein.
SQ SEQUENCE 1009 AA; 109834 MW; 80DD954F255C102B CRC64;

Query Match 21.4%; Score 820; DB 11; Length 1009;
Best Local Similarity 29.3%; Pred. No. 1.6e-52;
Matches 211; Conservative 109; Mismatches 261; Indels 138; Gaps 19;

Qy 1 EIKITRPSADGMLYNGQKRVPGSPPTLANRQDPFISGLVGGPRFFRFDAGSMATI 60
Db 410 QVTLFRAEADGLLLYCGESE-----HGRGDFMSLALIRSLHFRFCGTGIALI 460
Qy 61 RHPTPLALGHFTVTLRLSLITQSLIVGLDLPVNGTSQKFGQGLDNLNELYLGGYD-YG 119
Db 461 ISETKIKLGAHWITVLYRDGLNGMLQLNNGTPTVGOSQGYSKITFRTPLYLGGAPSAW 520
Qy 120 AIPKAGLSSGFCIGVRELRIQBEIVFHDNL-----LTAHGISHCPT--CRDRPCQNG 170
Db 521 LVRATGTNRGFCQVQSLSVNGKKI---DMRPWPLKALNGADVBCSSGICDEASCIHG 577
Qy 171 GQCHDESSYVVCVAGFTGSRCHESQALHCHPEACGPDATCVNRPDGRGYTCRHLGR 230
Db 578 GTCAAIKADSYICLPLGFRHCE-----DAPALTIPQ-----PR 613
Qy 231 SGURCEGVTVTPFSLSGAGSYLALPALTNTHHEL---RLDVEFKPLADPGVLLFS---G 284
Db 614 ESLR-----SYAATPWPLEPQHYLSFTEFEITFRDSDGCVLLIYSDTG 657
Qy 285 GKSGPVEDFVSLAMVGHLFRYELGSLAVLSAEPLALGRWHRVSAERLNKDGSLRVN 344
Db 658 SK-----DFLSINMAAGHVEFRFCGSGTGLRSEAPLTLGNWHLRVSTAKNGILQVD 712
Qy 345 GGRPVLSSPCKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLT 404
Db 713 KQVVEGMAEGGFTQIKCNTDIFIGVNPVDDVKNSGILHPFSGSIQKIIINDRTIHVK 772
Qy 405 YSFLGSGIGQCYDSSPCERQPCQHGATCMPAGEYEQICLRDGFKGDLCHEHENPCQLR 464
Db 773 HDF--TSGVNVENAAHPCVRAPCAHGSCRPKE-GYECDCPLGFEGELNCQKAI-----823
Qy 465 EPCILHGTCQGTRCILCPGSPRCQCGSGHGIAESDWHLESGGNDAPGQYGFHDDG 524
Db 824 -----IEAIEIPQFGRSYLTYPNLIK---RVSGSRN-----A 856
Qy 525 FLAPPHVFRSRLPEVPEETIEVLTSTASGILLWQGVGAGQKDFISLGLQDGHV 584
Db 857 FMR-----KTKADGLLNRG--DSPMRPNSDFISLGLRDGALI 894
Qy 585 FRYQLGSGEARLVSDEPINDGEWHRVTALREGRGSIQVDCGELYSRSPGNVAVNAKG 644
Db 895 PSYNLGSGVASIMVNGSFDNGRWHRYKAVRDGSGKITVDYGAETGKSPGLMRQLING 954
Qy 645 SVYIGGAPDVATLTGGRFSGSITGCVKNLVLSRSPGAPPQPLDLQHRAQAGANTRPC 703
Db 955 ALVYVGGMKEIALHTNRQYMRGLVGCISSHFTLST-----DVHISLVEDAVDGKNINTC 1006

RESULT 4
Q8NFS9 PRELIMINARY; PRT; 708 AA.
ID Q8NFS9
AC Q8NFS9;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
```



```
Query Match      21.2%: Score 812.5; DB 5; Length 1035;
Best Local Similarity 30.2%: Pred. No. 6e-52;
Matches 219; Conservative 121; Mismatches 304; Indels 81; Gaps 22;

QY 3 KITRPSADGMLYNGQKRVGSPNLANRQPPFISGLVGGPFRFRPAGSGMATIRH 62
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
370 ELTRPENGGLLLFNGCTRGSG-----DYIALSLKDRYAEFRPFGGKPMVLRA 419
QY 63 PTPALGHFTVTLLRSLTQGLSLVGLDLPVNGTSQGFQGLDNLNEELYLGGYDYGAI 122
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
420 EEPALNENHVRVSRPKRDYQIVDGHFVAFTLQOIPQLDIEDLYIGGVNWE LLP 479
QY 123 KAGLSS--GFTGCVRELRIQEEIVFHDNLNTANGISHCPTCRDRPONGGOCHDS-S 179
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
480 ADAVSQVQGVGCGISRLTQGRITVELIREAKYREGITDCRPAQGPQNGVCLSESQTE 539
QY 180 SYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCNRPDGRGYTCRCHLGRSLRCR 239
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
540 AYTCICQGWTRDCA-IEGTCQTPGVCGA--GRENTEEN--DMECFLFLNRSGRCCQYNE 595
QY 240 TVTTPSLSGAGSYLA---LPALTNTHTHELRLDVEFKPLA-PDGVLLFSFGKSGVDFVS 295
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
596 ILNHSILNFKNGSPAAFGTPKVT-----KVNITLSVRPASLEDSVILYTAESTLPSG 651
QY 296 LAMVGGHLEFRYELGSGI--AVLRSAPLALGRHVSAAERLNDKXGSLRVNGRPLRSS 353
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
652 LVIRGGHAELLINTAARLOPVVRSAPLPLNRWTRIEIRRLGEGILRVGDGPERKAKA 711
QY 354 POKSGLMLHTLLYLGGVEPS-VPLSPATNMSAHFRGCGVSVNGKRLDITYSFLSGOG 412
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
712 PGSDRIILSLKTHLYVGVYDRSVKVNDRDNTKPGDGCISRLYNFKPQVNLADIKDAAN 771
QY 413 IQCQVDSFPCRPQCHGATCMCPAGEYE-----FQCLCRDGFQKGLDCEHENPCQLREP 466
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
772 IQSC-----GETNMIGDESDNEPPVPPTPDVHNELOPYAMAF-C-ASDP 817
QY 467 CLHGTGCTQCTR-----CLCLPFGSGRCQGGSHGTAESDWHLESGNGNDAPGQYGFHD 522
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
818 CENGSGSCQEDVAVCSQFPGSGKHQCE-----HLQLG-----FNASFRG 858
QY 523 DGLAPPGHVFRSRLPEVPTTELEVRSTASGILLWGVGEAGQCKPISLGLDGH 582
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
859 DGYVNLNRSHFQPALEQSYTSIGVFTTKNGLLFWWGAGBEYTCQDFIAA VVDGY 918
QY 583 LVFRYQLSGEARLVSED--PINDGSHRVTLALREGRGSIQVDBGELVSGRS-PGPNAV 640
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
919 VYSYRLDGEAEVIRNSDIRVNGERHIVIAKRDENTALEVD-RMLHSGETRPTSKSM 977
QY 641 NAKGSYVIGGAPDVATLTGFRSSGITGCVKNVLHSAHPGAPPQPLDLOHRAQAGANT 700
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
978 KLPNGVFGGAPDLEVTGFRYKHLNGCI--VVVEGETVG-----QINTLSAAVNGVNA 1030
QY 701 RCPGS 705
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1031 NVCPA 1035

RESULT 6
Q8IRV9 PRELIMINARY; PRT; 4117 AA.
AC Q8IRV9;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE CG7981-PC.
GN TROL OR EG:BACR25B3.11 OR CG7981.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.H., Blazer B.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abrell J.F., Agbayani A., An H.J., Andrews-Frankoch C., Baldwin D.,
RA Beeson K.Y., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Casway S., Dahlke C., Davenport L.B., Davies P.,
RA de Pabloe B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodagat, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A.; Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Fargas V., Park S., Patel S., Pfeiffer B., Scheeler F.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak P., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
```


RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster";
RA Science 287:2185-2195(2000).
RN [2]
RN SEQUENCE FROM N.A.
RP Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houch J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirkas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RA "Sequencing of Drosophila melanogaster genome";
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RP Misra S., Crobey M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang X., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berner B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Seale S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RA "Annotation of Drosophila melanogaster genome";
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE FROM N.A.
RP Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RN SEQUENCE FROM N.A.
RP FlyBase;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF003424; AAF45786.3; -;
DR HSP; P00740; 1EDM.
DR FlyBase; FBgn0001402; trol.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR001525; C5_DNA_meth.
DR InterPro; IPR008985; ConA_like_rec_gl.
DR InterPro; IPR000742; EGF 2.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR000034; Laminin B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin G.
DR InterPro; IPR002172; LDL receptor_A.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00047; ig; 12.
DR Pfam; PF00052; laminin B; 3.
DR Pfam; PF00053; laminin_EGF; 2.
DR Pfam; PF00054; laminin_G; 3.
DR Pfam; PF00057; ldl_recept_a; 23.
DR PRINTS; PR00261; LDLRECEPTOR.

DR SMART; SM00180; EGF_Lam; 9.
DR SMART; SM00409; IG; 12.
DR SMART; SM00408; IGc2; 12.
DR SMART; SM00281; LamB; 3.
DR SMART; SM00282; LamG; 3.
DR SMART; SM00192; LDLA; 23.
DR PROSITE; PS00094; C5_MTASE_1; 1.
DR PROSITE; PS00022; EGF_1; 10.
DR PROSITE; PS01186; EGF_2; 6.
DR PROSITE; PS00835; IG_LIKE; 11.
DR PROSITE; PS01248; LAMININ TYPE EGF; 7.
DR PROSITE; PS00025; LAMININ TYPE EGF; 7.
DR PROSITE; PS01209; LDLRA_1; 20.
DR PROSITE; PS00068; LDLRA_2; 23.
DR PROSITE; PS00030; RRM_RNP_1; 1.
DR Immunoglobulin domain; Laminin EGF-like domain.
SQ SEQUENCE 4179 AA; 461782 MW; 47804277D0914E63 CRC64;
Query Match 21.2%; Score 812.5; DB 5; Length 4179;
Best Local Similarity 30.2%; Pred. No. 4e-51;
Matches 219; Conservative 121; Mismatches 304; Indels 81; Gaps 22;
QY 3 KITFPDSDAGMLLYNGQKRVPGSGPTNLNARQPDFISFGLVGRPEFRFDAGSGMATIRH 62
DB 3501 EUTFPENGDLLENGQTRSG-----DIALLSKDRIYAEFRDFGKPMVRA 3550
QY 63 PPTLALGHFTVTLRLSLTQGLIYVDLAPVNGTSQGRFQGLDLNLELYLGGYPYGAIP 122
DB 3551 EEPFALNENHVRVSRFKEDGYIQDEQHPVAPFTLQIIPQLDLIEDLYIGVGNWELLP 3610
QY 123 KAGLSS--GFICGVRELRLQGEIVFHDNLNLTAGHSICPTCEDPCONGGCHDSES-S 179
DB 3611 ADAVSQGVFGVGCISRLTLQGRTEVLEIRAKYKEGTLDCRCAQPCQNGKVCLESQTEQ 3670
QY 180 STVCVCPAGFTGSRCEHSQALCHPEACGPDATCVNRPDGRGYTCRCHLGRSLRCEGV 239
DB 3671 ATYCICQPGWTGRDCA-IGTQCTPGVCGA-CRCENTEN--DMECLCPLNRSGDRQYNE 3726
QY 240 TVTTPSLSAGSYLA---LPALTNTHTHELRLDVEKPLA-PGVLLFSGKSGPVEDFVS 295
DB 3727 ILNHSLSNFKGNSFAAYGTPKVT---KVNITLVRPASLEDSVILYTAESTLPSGDYLA 3782
QY 296 LAMVGHELFRYELGSL--AVLRSAEPLALGRVRSARLNKDGSLRVNGRPFVLRSS 353
DB 3783 LVLRGSHALLNTAARLDPVVRSAPPLNRTWTRIEIRRLGEGILRVGDGPERKAKA 3842
QY 354 PGKSGQLNHLTLLYIGGVFVS-VPLSPATNMSAHRGCVGVSNGKRLDLYSFILSGSG 412
DB 3843 PGSDRILSLKTHLYGVGYDRSTVKVNRDYNITKGFDCISRLYNFQKPNLLADIKDAAN 3902
QY 413 IGQCYDSSPCERQPCQHGATCMPAGEYE-----FOCLCRDGFKGDLCEHENPCQLREP 466
DB 3903 IQSC-----GETNMGDESDNEPPVPPPTPDVHENELOPYAMAPC-ASDP 3948
QY 467 CLHGGTCQGT-----CLCLFGSGPRCQSGSGHIAESDWHLEGGSGNDAPGQYGFHD 522
DB 3949 CENGSGCSEQEDVAVCVSCPFGSGKHCB-----HLQLG-----FNASFRG 3989
QY 523 DGFLAPGHVFSLSLEVPETIELEVTSTASGLLLWQGVGEAGQKDFISLGLQDGH 582
DB 3990 DGVVELNRSHFQPALEQSYTSMGIVTTKNKPLLFNWEQAGEEYTGQDFIAAVVDGY 4049
QY 583 LVFRYQLSGEARLVSED-PINDGEVHRTVALREGRRGSIQVDGELVSGRS-PGNVAV 640
DB 4050 VEYSMRLDGEAVIRNSDIRVDNGERHVIKAKRDENTAILVD-RMLHSGETFTSKKSM 4108
QY 641 NAGSVVYGGAPVAILTCGRFSSGITGCKNLVLHSAFPGAPPQPLDLOHRAQAGANT 700
DB 4109 KLPENFVVGAPDLEVTGFRYKHNLCICI--VVVEGETVG-----QINLSSAAVNGVNA 4161
QY 701 RCPFS 705
DB 4162 NVCPA 4166

RA Q8MPN3 PRELIMINARY; PRT; 4223 AA.
 AC Q8MPN3; TROL OR EG: BACR25B3.11 OR CG7981.
 DT 01-OCT-2002 (TRENDELrel. 22, Created)
 DT 01-OCT-2002 (TRENDELrel. 22, Last sequence update)
 DE Perlecan (CG7981-PD).
 GN TROL OR EG: BACR25B3.11 OR CG7981.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP Voigt A.;
 RT "Perlecan participates in proliferation activation of quiescent
 RT Drosophila neuroblasts".
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazer R.G., Champs M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballwe R.M., Basu A., Bakendell J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Borkan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
 RA De Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foele C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Markulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusekern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodgett J., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster".
 RL Science 287:2185-2195(2000).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP Celnik S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon K.Y., Beeson R.A., Busan D.A.,
 RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson K.A., Nelson K.A., Nunoo J.,
 RA Pacle J., Paragaa V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome".
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN SEQUENCE FROM N.A.
 RP Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celnik S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome".
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RN SEQUENCE FROM N.A.
 RP Adams M.D., Celnik S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RN SEQUENCE FROM N.A.
 RP FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ487018; CAD31650.1; -;
 DR EMBL; AE003424; AAN09079.1; -;
 DR FlyBase; FBgn0001402; trol.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0006306; P:DNA methylation; IEA.
 DR InterPro; IPRO01535; C5 DNA meth.
 DR InterPro; IPRO06985; CnA-like_rec_gl.
 DR InterPro; IPRO00742; EGF_2.
 DR InterPro; IPRO06209; EGF-like.
 DR InterPro; IPRO03593; IG.
 DR InterPro; IPRO07110; IG-like.
 DR InterPro; IPRO03598; IG_c2.
 DR InterPro; IPRO00034; Laminin_B.
 DR InterPro; IPRO02049; Laminin_EGF.
 DR InterPro; IPRO01791; Laminin_G.
 DR InterPro; IPRO02172; LDL_receptor_A.
 DR InterPro; IPRO00504; RNA_rec_mot.
 DR Pfam; PF00047; ig; 13.
 DR Pfam; PF00052; laminin_B; 3.
 DR Pfam; PF00053; laminin_EGF; 2.
 DR Pfam; PF00054; laminin_G; 3.
 DR Pfam; PF00057; ldl_recept_a; 23.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR ProDom; PD003031; Laminin_B; 3.
 DR SMART; SM00180; EGF_Lam; 8.
 DR SMART; SM00409; IG; 13.
 DR SMART; SM00408; IGC2; 13.
 DR SMART; SM00281; Lamb; 3.
 DR SMART; SM00282; Lamb; 3.
 DR SMART; SM00192; LDLA; 23.
 DR PROSITE; PS00094; C5_MTASE_1; 1.
 DR PROSITE; PS00022; EGF_1; 10.
 DR PROSITE; PS01186; EGF_2; 6.
 DR PROSITE; PS00835; IG LIKE; 12.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
 DR PROSITE; PS00225; LAM_G_DOMAIN; 3.
 DR PROSITE; PS01209; LDLRA_1; 20.
 DR PROSITE; PS00068; LDLRA_2; 23.
 DR PROSITE; PS00030; REM_RNP_1; 1.
 KW Immunoglobulin domain.
 SQ SEQUENCE 4223 AA; 456919 MW; 31D7C2B5C9B1D2E0 CRC64;

Query Match 21.2%; Score 812.5; DB 5; Length 4223;
 Best Local Similarity 30.2%; Pred. No. 4e-51;
 Matches 219; Conservative 121; Mismatches 304; Indels 81; Gaps 22;

Qy 333 ERLNKGSLRVNGRPRVLRSSPGKSGQLNHTLLYLGGVPSVFLSPATNMSAHPRGCVG 392
Db 173 ERNCRKALRVGDGPRVLGSPVHTVNLKEPLYVGADPFKSLARAAAASVSGFDGAIQ 232
Qy 393 EVSVNGKEL-----DLTVSFLSGIGQCYDDSSPCR---OPCORGATCMAGEY 439
Db 233 LVSGGGQQLITPEHVLKQVDVT-SFAG-----HPCIRASGHCPLNGASCVP---- 277
Qy 440 EFQCLCRDGFKGLDCEHEENPCQLREPLRGHTCGQTRCICLPFGSGPRCQOQSGHGIAE 499
Db 278 -----REAY-----VCLCPGFGSPHCEKGL----- 239
Qy 500 SDWHELGSGNDARPGQYCAFDGFLAFCHVFSRSLPVPET-----TELEVTS 551
Db 300 -----VEKSAG-----DVTAFDQRTFVNLNVTSEKALQNHLSIRTE 343
Qy 552 TASGLLWQGVGEAGCGKDFISLGLQDGHVFRYQLSGEARLVSEDPINGEWHVRT 611
Db 344 ATQGLVLS-----GKATERADYVALAIVDGHQLSYNLGSPVLRSTVPVNTNRLRVV 399
Qy 612 ALRGRGSGIOVDEELVSGSPGPNVAVNAKGSVYTGAPD--VALTGGFRSSGITGC 669
Db 400 AHREQREGSLQVNEAPVPTGSSPLGATQTDGALWLGGLPELVPGLPKAYGTGFGVC 459
Qy 670 VKNLVLSARPGAPPDPLDQHRQAQAGANTRPCPS 705
Db 460 LRDVVVGR-----HPLHLEDAVTKPELRPCPT 487
RESULT 11
Q9V714 PRELIMINARY; PRT: 1361 AA.
AC Q9V714;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE CG8403 protein.
GN SP2353 OR CG8403.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.N., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Stadlman A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Ebanon J., An H., Baldwin D., Barton J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Dou L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferriera S., Frise E., Galie R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Seattle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003808; AAF58071.2; -;
DR HSP; P08709; 1BF9.
DR FlyBase; FBgn0034070; SP2353.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00054; Laminin_G; 3.
DR SMART; SM00181; EGF; 4.
DR SMART; SM00282; LAMG; 3.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
KW EGF-like domain.
SQ SEQUENCE 1361 AA; 150298 MW; 4C71EDAB1A12ABED CRC64;
Query Match 17.4%; Score 664.5; DB 5; Length 1361;
Best Local Similarity 22.0%; Pred. No. 1e-40;
Matches 221; Conservative 111; Mismatches 295; Indels 379; Gaps 23;
Qy 1 EIKITPRDSADGMLNLYGKRVFGSPNTLANRQDPIFSLVGGPRPFRFDGSGNATI 60
Db 424 QLRTEPRFSPFDGIILLSGER-----DLDTGDFMALLNKGVFVFWFDCGSGVSV 474

QY 61 RHPTPLALGHFTVTLASLTQGLIVGLDAPVNGTSQKFGQGLDNEELYLGYPDY-G 119
Db 475 RSRETLINENWSVIYHRWDALVNLHGTVKQGRSNGLSRITFRFVPLGGIGNITG 534
QY 120 AIPKAGLSSGFGVRELRIOGEEIVFHDNLTAH-----GISHCPT--CRDRPCQ 168
Db 535 LAKRLPLAEGFAGCIR--RFVANE---HDYKFTHEPLGDVINGFDIQDSTDKCYRPPCQ 589
QY 169 NGQCHDSSESSVVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHL 228
Db 590 HGGKCLPDSQGA-ICLCPIGFVGLDCE----- 615
QY 229 GSGLRCEGVTVTTTSLSGAGSYLALPALNTNTHHELRLDVEPKPLADGVLLFSGKSG 288
Db 616 ----IRMD-----LQVPAFNGSSFLRYAPLGDLSALIWLELKVTLKPEQADGLIYSGPEHR 667
QY 289 PVDFVSLAMVGGHLEFRYELGSLAVLSAEPLALGRHVRSAERLNKDGSLRVNGGRP 348
Db 668 --GDFIALYNDGFVEFAFDLGGSPALVRSEHSLSGQWHTIKISRTARLAVLVKDKHQE 725
QY 349 VLRSFGKSGQGLNHLTLILYLGVEPSVPLSPATNMSAHFRGCVGEVSNKRLDLYSFL 408
Db 726 VLTISNGFWHLSDQNLFGVGNVHVDRLFLDLKYKFFVFGCIQRIDINGHSLGIVAEAL 785
QY 409 GSGQIGQCVSSPCEROPCHGATCMPAGE-YEFQCLCRDGFKGLDCEH--EENPCQLRE 465
Db 786 GGSNIGNC--PHACVARPCGPLAECVPQWESYE----CRCSIHNERCNKAAEVPEQLPE 839
QY 466 PCLHGTCQGT----- 477
Db 840 LALHKSQVLETKDNGEAAKVSGLAHKSKHRLNHLKPTATSTTTTSTTTTTEAP 899
QY 478 -----CLCL 481
Db 900 SERTEATAGALSNEEIEDIIIFRLVQOQOQKELKKHQOQTTTATATSTSSGSKAK 959
QY 482 PGFSG----- 486
Db 960 PRLSGKHASKHEHHLKPNAAFRLKSLRPLTHYESFQTNPDSDILTFEDNNDWVTSLOOQ 1019
QY 487 -----PRCQOQ----- 492
Db 1020 EYGDAAASQVPLAFEDASGFTFRSDNEDDENAFVDESIFDASDGTTEEYQKQLAQD 1079
QY 493 -----SGHGAESDMLHGS-----GNDAPGQYGAVFHDDGFL----- 526
Db 1080 MKRIMSNASHSHKKAQVQFPQSGQEVGTANEDTSQYSDDYNDDELLTPVMQGBEVK 1139
QY 527 -----APPG-----HV 532
Db 1140 LEQHTSTPTQTHDWSLLKFDLSAHSQVQGVKRNKFCAGFAGSDSYFHYNDATMSQV 1199
QY 533 FSRSLPEVETIEVTRTASGLLQGVGEAGQKDFISLGLQDGLHVFYQLQSG 592
Db 1200 ISYSL-----DLNLRKTHSENGVILWTGRQ-GTTEHDDYLSLGTQGYLHFRYDLGSG 1253
QY 593 EARL-VSEDPINDGEHVRVTLAREGRSTQVTDGEELVSGRPGPNVAVNAGSVYVIGGA 651
Db 1254 EVDIREFNGTKVSDGLHVRVRAINSQEGYLEVDGRKTVTLRPAKGLRQNTDTGLYVGGM 1313
QY 652 PDVATLTGGRFSSGITGCVNVLHSAHPGAPPPQPLDLQRAQAG 697
Db 1314 PDVGYFTHQRYFSGVIGCISEIVLAGEMKLNFDNTLGTTEHNVETG 1359
RESULT 12
Q9NGV2
ID Q9NGV2 PRELIMINARY; PRT; 1361 AA.
AC Q9NGV2
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE SP2353.
GN Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Serano T.L., Pendleton J.D., Rubin G.M.;
RT "A reverse genetic screen for genes involved in Drosophila
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF239610; AAF63502.1; -.
DR HSP; P08709; 1BF9.
DR FlyBase; FBgn0034070; SP2353.
DR InterPro; IPR008985; ConA like_1ec_g1.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR002210; IEGF.
DR InterPro; IPR001791; Laminin G.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00054; laminin G; 3.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00282; LamG; 3.
DR PROSITE; PS00022; EGF 1; 2.
DR PROSITE; PS01186; EGF 2; 2.
DR PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW EGF-like domain.
SQ SEQUENCE 1361 AA; 150355 MW; 206D9F9BF9EDFE47 CRC64;
Query Match 17.4%; Score 664.5; DB 5; Length 1361;
Best Local Similarity 22.0%; Pred. No. 1e-40;
Matches 221; Conservative 111; Mismatches 295; Indels 379; Gaps 23;
QY 1 EIKITRPSADQMLYNGKRVGSPFTNANRQPDFISGLVGGPFRFDAGSGMATI 60
Db 424 QLRIFRPSDFDGIILLGSR-----DDLTGDFMALLNKGVFWFDCGSGVSV 474
QY 61 RHPTPLALGHFTVTLRLSLTQGLIVGLDAPVNGTSQKFGQGLDNEELYLGYPDY-G 119
Db 475 RSRETLINENWSVIYHRWDALVNLHGTVKQGRSNGLSRITFRFVPLGGIGNITG 534
QY 120 AIPKAGLSSGFGVRELRIOGEEIVFHDNLTAH-----GISHCPT--CRDRPCQ 168
Db 535 LAKRLPLAEGFAGCIR--RFVANE---HDYKFTHEPLGDVINGFDIQDSTDKCYRPPCQ 589
QY 169 NGQCHDSSESSVVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHL 228
Db 590 HGGKCLPDSQGA-ICLCPIGFVGLDCE----- 615
QY 229 GSGLRCEGVTVTTTSLSGAGSYLALPALNTNTHHELRLDVEPKPLADGVLLFSGKSG 288
Db 616 ----IRMD-----LQVPAFNGSSFLRYAPLGDLSALIWLELKVTLKPEQADGLIYSGPEHR 667
QY 289 PVDFVSLAMVGGHLEFRYELGSLAVLSAEPLALGRHVRSAERLNKDGSLRVNGGRP 348
Db 668 --GDFIALYNDGFVEFAFDLGGSPALVRSEHSLSGQWHTIKISRTARLAVLVKDKHQE 725
QY 349 VLRSFGKSGQGLNHLTLILYLGVEPSVPLSPATNMSAHFRGCVGEVSNKRLDLYSFL 408
Db 726 VLTISNGFWHLSDQNLFGVGNVHVDRLFLDLKYKFFVFGCIQRIDINGHSLGIVAEAL 785
QY 409 GSGQIGQCVSSPCEROPCHGATCMPAGE-YEFQCLCRDGFKGLDCEH--EENPCQLRE 465
Db 786 GGSNIGNC--PHACVARPCGPLAECVPQWESYE----CRCSIHNERCNKAAEVPEQLPE 839
QY 466 PCLHGTCQGT----- 477
Db 840 LALHKSQVLETKDNGEAAKVSGLAHKSKHRLNHLKPTATSTTTTSTTTTTEAP 899
QY 478 -----CLCL 481
Db 900 SERTEATAGALSNEEIEDIIIFRLVQOQOQKELKKHQOQTTTATATSTSSGSKAK 959


```

QY 482 PGFSG----- 486
Db 960 PRLSGKHASKHEHLKPNAAFRKLRLPLTHYESFQTNFDSILTLTFEDNNDWVTSLOQQ 1019
QY 487 -----PRCOQG----- 492
Db 1020 EYGDMAAASQVPLAFEDASPGTPRSDNNEEDENAFVDESIFDASDGTVEYORKQLAQD 1079
QY 493 -----SGHIAESDMHLEGS-----GGNDAPGQYAGFYHDDGFL----- 526
Db 1080 MKRIMSNAHSHKKAQVQFPQGSQEVGTANEDTSQYSDYNDDELITPVMQGGEEVK 1139
QY 527 -----APFG-----HV 532
Db 1140 LEQHTSTPQTHDWSLLKFLDSAHQSQVQVRKQFACFAGSDSYHYNDADTMSQV 1199
QY 533 FRSLSPEVPEETIELEVRTSTASGLLWQVGEAGQKDFISLQDGLHVLFRYQLGSG 592
Db 1200 ISYSI-----DLNLRKTHSENGVILWTGRQ-OTTEHDDYLSLGIEQGYLHFRYDLGSG 1253
QY 593 EARL-VSEDPINDEGWHRVLTALREGRGSIQVDEBELVSGRSPGPNVAVNAKGSVYIGGA 651
Db 1254 EVDIRFNGTKVSDGLHVRVRAIRNSQSVLEVDGRKTVTLRAPKLRQLNTDTGLYVGM 1313
QY 652 PDVATUTGGRFSGITGCVNVLHSAKPGAPPPQPLDLQHRAQAG 697
Db 1314 PDVGYFTHQYFSGIVGCISEIVLAGEMKLNFPDNTLGTENHVEIG 1359

RESULT 13
Q8NAL2 PRELIMINARY; PRT; 463 AA.
ID Q8NAL2
AC Q8NAL2;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ35160.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
TISSUE=Placenta;
RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isegai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK092479; BAC03900.1; -.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00054; laminin G; 2.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00179; EGF_Ca; 2.
DR SMART; SM00282; Lact; 2.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS00025; LAM_G DOMAIN; 2.
KW Hypothetical protein; EGF-like domain.
SQ SEQUENCE 463 AA; 50635 MW; F6889F2714D5D0EC CRC64;

```

Query Match 16.4%; Score 628; DB 4; Length 463;
 Best Local Similarity 28.7%; Pred. No. 1.3e-38;
 Matches 162; Conservative 80; Mismatches 198; Indels 124; Gaps 16;

```

QY 151 LTAHGISHCPT--CRDRPQNGQCHDSSESSVYVCPAGCTGSRCHESQALHCHPEACG 208
Db 10 LSGADVGECSGICDEASCIHGGTCTA KADSVICLCPLGF----- 50
QY 209 PDATCVNRPDGRGYTCRCHLGRSLRCBEGVTWTPSL-SGAGSYLALPALTNTHHEL-- 265
Db 51 -----KGRHCEDAFTLTIPOFRESLSYAATPWPLEPQHLYSF 88
QY 266 -RLDVEFKPLADQVLLPS---GKSGPVEDFVSLAMVGGHLEPRYELGSLAVRSEAP 321
Db 89 MEFEITPRPDSGDGLLYSYDTGSK-----DFLSINLAGHVEFRPCGSGTGVLRSEDP 143
QY 322 LALGRWVRVSAERLNKDGSLRVNGRPLVRSSPGKSQGLNLHTLLYLGVEPSVPLSPAT 381
Db 144 LTLGNWHELVRSTAKNGILLQVDKQKIVEGMAEGFTQIKCNTDIFIGGVNVDVKNS 203
QY 382 NMSAHFGCVGEVSVNGKRLDLTVSFLGSGIGOCYDSSPCEROPCOHGATCMPAGEYEF 441
Db 204 GVLKPPFSGSIQKIILNDRTHVHRDF--TSGVNVENAAHPCVRAPCAHGGSCRPKE-GY 260
QY 442 QCLCRDGFKGDLCEHENPQOLREPCLHGGTCCQT--RCLCLPGFSGPRCQOQSGHGIAE 499
Db 261 DCDPLGPEG-----LH---CQKAIIEAIEIPQFIG----- 288
QY 500 SDMLHLEGGNDAPGQYAGFYHDDGFLAFPGHVFPSRLPEVPEETIELEVRTSTASGLLW 559
Db 289 -----RSYLTVDNPDILKRVSG---SRS-----NVFMRFKTKAKDGLLLW 325
QY 560 QGVEVGEAGQKDFISLQDGLHVFYVQLGSGEARLVSEDPINDGEWHRVTLREGRRG 619
Db 326 RG--DSPMRPNSDFISLGRDGLVFSYNLGSGVASIMVNGSFNDGRHVRKAVRDQSG 383
QY 620 SIQVDGEELVSGRPGPNVAVNAKGSVYIGAPDVATLTGGRFPSSGITGVKNLVLHAR 679
Db 384 KITVDYDYGARTKSPGMMRQLNGLNGALYVGGMKBIALHTNRQYMRGLVGGCIHSHTLST-- 441
QY 680 PGAPPPQPLDLQHRAQAGANTRPC 703
Db 442 -----DYHISLVEDAVDGKNINTC 460

RESULT 14
Q8BU56 PRELIMINARY; PRT; 822 AA.
ID Q8BU56
AC Q8BU56;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical EGF-like domain.
GN AU040377.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
STRAIN=C57BL/6J; TISSUE=Oviduct;
RC MEDLINE=23254683; PubMed=12466851;
RX The PANTOM Consortium,
RA the Riken Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573 (2002).
RL EMBL; AK087659; BAC39956.1; -.
DR MGD; MGI:2146149; AU040377.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR008985; Cona-like_lec_g.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR003962; FnlII subd.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR008957; FN_III-like.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:19:48 ; Search time 85.8465 seconds
(without alignments)
2320.373 Million cell updates/sec

Title: US-10-006-011A-3
Perfect score: 705
Sequence: 1 EKLTRPDSADGMLLYNQ.....QPLDQHRAGAGNTRPCPS 705

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseq1980a.*
2: Geneseq1990a.*
3: Geneseq2000a.*
4: Geneseq2001a.*
5: Geneseq2002a.*
6: Geneseq2003a.*
7: Geneseq2003bs.*
8: Geneseq2004a.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	705	100.0	4391	6	Aae34390 Human per
2	443	62.8	4393	4	Aab31889 Amino aci
3	383	54.3	4436	4	Abg23265 Novel hum
4	195	27.7	195	4	Aab31890 Amino aci
5	18	2.6	18	4	Abb55911 Vascular
6	14	2.0	14	4	Abb55908 Vascular
7	14	2.0	14	4	Abb55905 Vascular
8	10	1.4	10	4	Abb55910 Vascular
9	9	1.3	15	4	Abb55909 Vascular
10	9	1.3	159	6	Abm65732 Propionib
11	9	1.3	2531	7	Adg63713 Rat Prote
12	9	1.3	2531	7	Adg63705 Rat Prote
13	9	1.3	2531	7	Adg63709 Rat Prote
14	9	1.3	2531	7	Adg63701 Rat Prote
15	9	1.3	3319	4	Abb70376 Drosophil
16	8	1.1	8	4	Abb55907 Vascular
17	8	1.1	14	4	Abb55906 Vascular
18	8	1.1	36	5	Aae13652 Mouse cyc
19	8	1.1	47	5	Aae13653 Human e-s
20	8	1.1	54	4	Aau40544 Propionib
21	8	1.1	54	6	Abm37063 Propionib
22	8	1.1	54	7	Abg23342 Human sec
23	8	1.1	55	2	Aay01436 Secreted
24	8	1.1	66	5	Aae13647 Human fib
25	8	1.1	77	4	Aau63351 Propionib

26	8	1.1	77	6	Abm59870 Propionib
27	8	1.1	80	4	Aam15081 Peptide #
28	8	1.1	80	4	Abb35070 Peptide #
29	8	1.1	80	4	Aam28574 Peptide #
30	8	1.1	80	4	Abb29890 Peptide #
31	8	1.1	80	4	Abb20486 Protein #
32	8	1.1	80	4	Aam68258 Human bon
33	8	1.1	80	4	Aam55885 Human bra
34	8	1.1	80	4	Abg49912 Human liv
35	8	1.1	80	4	Aam03808 Peptide #
36	8	1.1	80	5	Abg37793 Human pep
37	8	1.1	82	5	Aae13646 Human coa
38	8	1.1	83	5	Aae13645 Pig facto
39	8	1.1	88	5	Aae13648 Human coa
40	8	1.1	110	5	Abb53124 Human ORF
41	8	1.1	123	5	Aae13650 Human ins
42	8	1.1	131	5	Abb90171 Human pol
43	8	1.1	143	5	Aae13643 Mouse lam
44	8	1.1	144	5	Abb53123 Human ORF
45	8	1.1	144	5	Abb53141 Human ORF

ALIGNMENTS

RESULT 1
ID AAE34390 standard; protein; 4391 AA.
XX
AC AAE34390;
XX
DT 14-MAY-2003 (first entry)
XX
DE Human perlecan protein.
XX
KW Human; diagnosis; osteoarthritis; rheumatoid arthritis; perlecan.
XX
OS Homo sapiens.
XX
FN WO200295415-A2.
XX
PD 28-NOV-2002.
XX
PF 22-MAY-2002; 2002WO-EP005612.
XX
PR 23-MAY-2001; 2001GB-00012626.
XX
PA (OSTE-) OSTEOMETER BIO TECH AS.
XX
PI Christgau S, Henriksen DB, Cloos PAC;
XX
DR WPI; 2003-140389/13.
XX
PT An assay for the diagnosis or assessment of the severity of
PT osteoarthritis or rheumatoid arthritis comprising detecting an isomerized
PT or optically inverted protein in a sample.
XX
PS Disclosure; Page 46-67; 106pp; English.
XX
CC The invention relates to an assay for the diagnosis or assessment of the
CC severity of osteoarthritis or rheumatoid arthritis. The assay involves
CC measuring (in a biological sample) the amount or presence of an
CC isomerized or optically inverted protein or one or more isomerized or
CC optically inverted fragments from proteins such as perlecan, biglycan,
CC decorin, fibrillin-1 or protocadherin. The assay is useful for the
CC diagnosis or assessment of the severity of osteoarthritis or rheumatoid
CC arthritis. The present sequence is human perlecan protein
XX
SQ Sequence 4391 AA;

Query Match 100.0%; Score 705; DB 6; Length 4391;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EIKITRPSADGMLYNGKRVPGSPTNLANKRDPFISFGLVGRPEFRFDAGSGWATI 60
 DB 3687 EIKITRPSADGMLYNGKRVPGSPTNLANKRDPFISFGLVGRPEFRFDAGSGWATI 3746
 QY 61 RHTPLALGHFHTVTLRLSLTQSLIVGLDAPVNGTSQKFGQGLDNEELYLGYPYGA 120
 DB 3747 RHTPLALGHFHTVTLRLSLTQSLIVGLDAPVNGTSQKFGQGLDNEELYLGYPYGA 3806
 QY 121 IPKAGSSSGIGCVRELRIQGEIEIVFHDNLTAHGISHCPTCRDRPCQNGQCHDSSESS 180
 DB 3807 IPKAGSSSGIGCVRELRIQGEIEIVFHDNLTAHGISHCPTCRDRPCQNGQCHDSSESS 3866
 QY 181 YVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEGVT 240
 DB 3867 YVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEGVT 3926
 QY 241 VTPPSLGSAGSYLALPALTNTNTHHELDVFEKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
 DB 3927 VTPPSLGSAGSYLALPALTNTNTHHELDVFEKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 3986
 QY 301 GHLEFRYELGSLAVLRSAPLALGRHVRSAERLNKDGSLRVNGRPLVRSFGKSGQL 360
 DB 3987 GHLEFRYELGSLAVLRSAPLALGRHVRSAERLNKDGSLRVNGRPLVRSFGKSGQL 4046
 QY 361 NLHTLLYLGGVEPSVPLSPATNMSAHRFCVGEVSVNGKRLDLYSFLGSGQIGQCYDSS 420
 DB 4047 NLHTLLYLGGVEPSVPLSPATNMSAHRFCVGEVSVNGKRLDLYSFLGSGQIGQCYDSS 4106
 QY 421 PCERQPCQHGATCMPAGEYEFQCLCRDGIKGDLCHEENPCQLRHPCLHGGTCQGRCLC 480
 DB 4107 PCERQPCQHGATCMPAGEYEFQCLCRDGIKGDLCHEENPCQLRHPCLHGGTCQGRCLC 4166
 QY 481 LPFGSGPRCQCGSGHGAIESDWHLEGGGNDAPQCYGAYFHDGFLAPFGHVSRSIPEV 540
 DB 4167 LPFGSGPRCQCGSGHGAIESDWHLEGGGNDAPQCYGAYFHDGFLAPFGHVSRSIPEV 4226
 QY 541 PETIELEVRTSTAGLLWQGVGEVAGGQKDFISLIGDQHLVFRYQLGSGEARLYSED 600
 DB 4227 PETIELEVRTSTAGLLWQGVGEVAGGQKDFISLIGDQHLVFRYQLGSGEARLYSED 4286
 QY 601 PINDGEHVRTALREGRGSIQVDGELVSGRSPGNVAVNAKGSVYTGAPDVATLTGG 660
 DB 4287 PINDGEHVRTALREGRGSIQVDGELVSGRSPGNVAVNAKGSVYTGAPDVATLTGG 4346
 QY 661 RFSSGITGVKNLVLSARPCAPPDPLDIQHRQAQANTRECPSS 705
 DB 4347 RFSSGITGVKNLVLSARPCAPPDPLDIQHRQAQANTRECPSS 4391

RESULT 2
 AAB31889
 ID AAB31889 standard; protein; 4393 AA.
 XX AC AAB31889;
 XX AC AAB31889;
 DT 15-MAY-2001 (first entry)
 DE Amino acid sequence of a human protein.
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW Ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

OS Homo sapiens.
 XX FN WO200105422-A2.
 XX PD 25-JAN-2001.
 XX PF 17-JUL-2000; 2000WO-FR002057.

XX 15-JUL-1999; 99FR-00009372.
 XX (INNR) BIOMERIEUX STELHYS.
 XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX WPI; 2001-159475/16.
 XX Detecting, preventing and treating degenerative, neurological and
 XX autoimmune diseases, particularly multiple sclerosis, using specified
 XX polypeptides or related nucleic acid or ligand.
 XX Claim 1; Page 138-152; 209pp; French.
 XX The present sequence represents a human protein, which is used in the
 XX method of the invention. The specification describes a method which uses
 XX at least one polypeptide or polynucleotide sequence belonging to the
 XX perlecan, precursor of the retinol-binding plasma protein, precursor of
 XX the ganglioside GM2 activator, calgranulin B or saposin B protein
 XX families. The method is used for detecting, preventing or treating a
 XX degenerative, neurological and/or auto-immune disease. The
 XX polynucleotides and polypeptides are used for diagnosis, prognosis, and
 XX prevention and treatment of multiple sclerosis (in its various forms and
 XX phases). They may also be useful in cases of e.g. Alzheimer's and
 XX Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 XX polyarthritis and lupus erythematosus, including use as vaccines and in
 XX gene therapy (expression of sense or antisense sequences). They can also
 XX be used to assess efficacy of potential therapeutic agents, particularly
 XX compounds that reduce or inhibit toxicity towards glial cells

Sequence 4393 AA;

Query Match 62.8%; Score 443; DB 4; Length 4393;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 643; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EIKITRPSADGMLYNGKRVPGSPTNLANKRDPFISFGLVGRPEFRFDAGSGWATI 60
 DB 3689 EIKITRPSADGMLYNGKRVPGSPTNLANKRDPFISFGLVGRPEFRFDAGSGWATI 3748
 QY 61 RHTPLALGHFHTVTLRLSLTQSLIVGLDAPVNGTSQKFGQGLDNEELYLGYPYGA 120
 DB 3749 RHTPLALGHFHTVTLRLSLTQSLIVGLDAPVNGTSQKFGQGLDNEELYLGYPYGA 3808
 QY 121 IPKAGSSSGIGCVRELRIQGEIEIVFHDNLTAHGISHCPTCRDRPCQNGQCHDSSESS 180
 DB 3809 IPKAGSSSGIGCVRELRIQGEIEIVFHDNLTAHGISHCPTCRDRPCQNGQCHDSSESS 3868
 QY 181 YVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEGVT 240
 DB 3869 YVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEGVT 3928
 QY 241 VTPPSLGSAGSYLALPALTNTNTHHELDVFEKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
 DB 3929 VTPPSLGSAGSYLALPALTNTNTHHELDVFEKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 3988
 QY 301 GHLEFRYELGSLAVLRSAPLALGRHVRSAERLNKDGSLRVNGRPLVRSFGKSGQL 360
 DB 3989 GHLEFRYELGSLAVLRSAPLALGRHVRSAERLNKDGSLRVNGRPLVRSFGKSGQL 4048
 QY 361 NLHTLLYLGGVEPSVPLSPATNMSAHRFCVGEVSVNGKRLDLYSFLGSGQIGQCYDSS 420
 DB 4049 NLHTLLYLGGVEPSVPLSPATNMSAHRFCVGEVSVNGKRLDLYSFLGSGQIGQCYDSS 4108
 QY 421 PCERQPCQHGATCMPAGEYEFQCLCRDGIKGDLCHEENPCQLRHPCLHGGTCQGRCLC 480
 DB 4109 PCERQPCQHGATCMPAGEYEFQCLCRDGIKGDLCHEENPCQLRHPCLHGGTCQGRCLC 4168
 QY 481 LPFGSGPRCQCGSGHGAIESDWHLEGGGNDAPQCYGAYFHDGFLAPFGHVSRSIPEV 540
 DB 4169 LPFGSGPRCQCGSGHGAIESDWHLEGGGNDAPQCYGAYFHDGFLAPFGHVSRSIPEV 4228

QY 541 PETIELEVRTSTASGLLLWQGVGEAGQKDFISLGLQDGHVFRYQLGSGEARLYSED 600
DB 4229 PETIELEVRTSTASGLLLWQGVGEAGQKDFISLGLQDGHVFRYQLGSGEARLYSED 4288
QY 601 PINDGEHVRVTLRGGRGSIQVDGEELVSGRSPGNVAVNAKGS 645
DB 4289 PINDGEHVRVTLRGGRGSIQVDGEELVSGRSPGNVAVNAKGS 4333

RESULT 3
ABG23265
ID ABG23265 standard; protein; 4436 AA.
XX

AC ABG23265;
DT 18-FEB-2002 (first entry)
XX

DE Novel human diagnostic protein #23256.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX WO200175067-A2.

PN 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSZ-) HYSEQ INC.

XX Dmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS87452.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX Claim 20; SEQ ID NO 53624; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 4436 AA;

Query Match 54.3%; Score 383; DB 4; Length 4436;
Best Local Similarity 99.7%; Pred. No. 0;

Matches 583; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRPDPFISGLVGGPFRFADAGSMATI 60
DB 3722 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRPDPFISGLVGGPFRFADAGSMATI 3781
QY 61 RHPTPLALGHFTVTLRLSITQSLIVGDLAPVNGTSQKFGQLDNEELYLGYPDYGA 120
DB 3782 RHPTPLALGHFTVTLRLSITQSLIVGDLAPVNGTSQKFGQLDNEELYLGYPDYGA 3841
QY 121 IPKAGLSSGFIGVCVRELRIOGEBIVFHDLMNLTAHGISHCPTCRDRPCQNGQCHDSSESS 180
DB 3842 IPKAGLSSGFIGVCVRELRIOGEBIVFHDLMNLTAHGISHCPTCRDRPCQNGQCHDSSESS 3901
QY 181 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCNRPDGRGYTCRCHLGRSGLRCEGVT 240
DB 3902 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCNRPDGRGYTCRCHLGRSGLRCEGVT 3961
QY 241 VTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGKSGPVEDFVSLAMVG 300
DB 3962 VTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGKSGPVEDFVSLAMVG 4021
QY 301 CHLEFRYELSGSLAVLSAEPLALGRWHRYSAERLNKDGSLRVNGGPPVLRSSPGKSQGL 360
DB 4022 CHLEFRYELSGSLAVLSAEPLALGRWHRYSAERLNKDGSLRVNGGPPVLRSSPGKSQGL 4081
QY 361 NLHTLLYLGVEPSVPLSPATNMSAHPRGCVGEVSVNGKRLDLYTSFLSGQIGQCYDSS 420
DB 4082 NLHTLLYLGVEPSVPLSPATNMSAHPRGCVGEVSVNGKRLDLYTSFLSGQIGQCYDSS 4141
QY 421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCOLREPCCLHGTCCGTCLC 480
DB 4142 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCOLREPCCLHGTCCGTCLC 4201
QY 481 LPFGSGPRCQCGSGHGAESDWHLEGGSGNDAPQCYGAYFHDGGLAFPPGHVFSRSLPEV 540
DB 4202 LPFGSGPRCQCGSGHGAESDWHLEGGSGNDAPQCYGAYFHDGGLAFPPGHVFSRSLPEV 4261
QY 541 PETIELEVRTSTASGLLLWQGVGEAGQKDFISLGLQDGHVFRYQLGSGEARLYSED 585
DB 4262 PETIELEVRTSTASGLLLWQGVGEAGQKDFISLGLQDGHVFRYQLGSGEARLYSED 4306

RESULT 4

AAB31890

ID AAB31890 standard; protein; 195 AA.

XX AAB31890;

XX 15-MAY-2001 (first entry)

DE Amino acid sequence of the C-terminal of the human perlecan protein.

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

XX Homo sapiens.

XX WO200105422-A2.

XX 25-JAN-2001.

XX 17-JUL-2000; 2000WO-FR002057.

XX 15-JUL-1999; 99FR-00009372.

XX (INMR) BIOMERIEUX STELHYS.

XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;

XX

DR WPI; 2001-159475/16.
 DR N-PSDB; AAF54728.
 XX
 PT Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 XX Claim 1; Page 152-153; 209pp; French.
 XX
 CC The present sequence represents a human polypeptide, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthrititis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 XX Sequence 195 AA;
 SQ
 Query Match 27.7%; Score 195; DB 4; Length 195;
 Best Local Similarity 100.0%; Pred. No. 5.3e-183; Indels 0; Gaps 0;
 Matches 195; Conservative 0; Mismatches 0;
 QY 511 DAPGOYAYFHDGFLAPFHVFSRSLPEVETIELEVRTSTASGLLLWQGVGEAGQG 570
 Db 1 DAPGOYAYFHDGFLAPFHVFSRSLPEVETIELEVRTSTASGLLLWQGVGEAGQG 60
 QY 571 KDFISLGLQDGHVFRYOLGSGEARLVSEDPINDGEWHRVTLREGRGSLQVDGELVS 630
 Db 61 KDFISLGLQDGHVFRYOLGSGEARLVSEDPINDGEWHRVTLREGRGSLQVDGELVS 120
 QY 631 GRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLSHARFGAPPPQPLDL 690
 Db 121 GRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLSHARFGAPPPQPLDL 180
 QY 691 QHRAQAGANTRPCPS 705
 Db 181 QHRAQAGANTRPCPS 195
 RESULT 5
 ABB55911
 ID ABB55911 standard; peptide; 18 AA.
 XX
 AC ABB55911;
 XX
 DT 15-FEB-2002 (first entry)
 XX
 DE Vascular dementia-associated protein isoform (VPI) 111.
 XX
 KW Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
 KW diagnosis; prognosis; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200169261-A2.
 XX
 XX 20-SEP-2001.
 XX
 PF 14-MAR-2001; 2001WO-GB001106.
 XX
 PR 15-MAR-2000; 2000GB-00006285.
 PR 24-NOV-2000; 2000GB-00028734.
 PR 28-NOV-2000; 2000US-00724391.
 XX
 XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
 XX
 XX Herath HMAc, Parekh RB, Rohlf C;
 XX
 XX WPI; 2001-557937/62.
 XX
 XX Screening, diagnosis or prognosis of vascular dementia (VD), useful for
 XX determining stage of VD and monitoring the effect of VD therapy,
 XX comprises analyzing body fluid by 2-dimensional electrophoresis for
 XX features correlated with VD.

(OXFO-) OXFORD GLYCOSCIENCES UK LTD.
 PA Herath HMAc, Parekh RB, Rohlf C;
 XX
 XX WPI; 2001-557937/62.
 XX
 PT Screening, diagnosis or prognosis of vascular dementia (VD), useful for
 PT determining stage of VD and monitoring the effect of VD therapy,
 PT comprises analyzing body fluid by 2-dimensional electrophoresis for
 PT features correlated with VD.
 XX
 XX Claim 6; Page 32; 151pp; English.
 XX
 CC The invention relates to screening, diagnosis or prognosis of Vascular
 CC Dementia (VD) in a subject comprising analysing body fluid from the
 CC subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
 CC features containing at least one chosen feature whose relative abundance
 CC correlates with the presence, absence, stage or severity of VD or
 CC predicts the onset or course of VD, especially detecting in a sample of
 CC cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
 CC protein isoforms (VPIs) (ABB55901-ABB56295) as fully defined in the
 CC specification. Detecting VD-associated features and VPI is useful for the
 CC screening, diagnosis or prognosis of VD, for determining the stage or
 CC severity of VD, for identifying a subject at risk of VD or for monitoring
 CC the effect of therapy administered to a subject having VD. Nucleic acids
 CC encoding a VPI or inhibiting the function of a VPI are useful for the
 CC treatment of VD and for gene therapy
 XX
 XX Sequence 18 AA;
 SQ
 Query Match 2.6%; Score 18; DB 4; Length 18;
 Best Local Similarity 100.0%; Pred. No. 9.5e-10;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 644 GSVYIGGAPDVATLTGGR 661
 Db 1 GSVYIGGAPDVATLTGGR 18
 RESULT 6
 ABB55908
 ID ABB55908 standard; peptide; 14 AA.
 XX
 AC ABB55908;
 XX
 DT 15-FEB-2002 (first entry)
 XX
 DE Vascular dementia-associated protein isoform (VPI) 108.
 XX
 KW Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
 KW diagnosis; prognosis; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200169261-A2.
 XX
 XX 20-SEP-2001.
 XX
 PF 14-MAR-2001; 2001WO-GB001106.
 XX
 PR 15-MAR-2000; 2000GB-00006285.
 PR 24-NOV-2000; 2000GB-00028734.
 PR 28-NOV-2000; 2000US-00724391.
 XX
 XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
 XX
 XX Herath HMAc, Parekh RB, Rohlf C;
 XX
 XX WPI; 2001-557937/62.
 XX
 XX Screening, diagnosis or prognosis of vascular dementia (VD), useful for
 XX determining stage of VD and monitoring the effect of VD therapy,
 XX PT comprises analyzing body fluid by 2-dimensional electrophoresis for
 XX PT

PT features correlated with VD.
 XX Claim 6; Page 32; 151pp; English.
 XX The invention relates to screening, diagnosis or prognosis of Vascular
 CC Dementia (VD) in a subject comprising analysing body fluid from the
 CC subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
 CC features containing at least one chosen feature whose relative abundance
 CC correlates with the presence, absence, stage or severity of VD or
 CC predicts the onset or course of VD, especially detecting in a sample of
 CC cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
 CC protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
 CC specification. Detecting VD-associated features and VPI is useful for the
 CC screening, diagnosis or prognosis of VD, for determining the stage or
 CC severity of VD, for identifying a subject at risk of VD or for monitoring
 CC the effect of therapy administered to a subject having VD. Nucleic acids
 CC encoding a VPI or inhibiting the function of a VPI are useful for the
 CC treatment of VD and for gene therapy
 XX Sequence 14 AA;
 SQ

Query Match 2.0%; Score 14; DB 4; Length 14;
 Best Local Similarity 100.0%; Pred. No. 6.5e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 596 LVSEDPINDGEWHR 609
 DB 1 LVSEDPINDGEWHR 14

RESULT 7
 ABB55905
 ID ABB55905 standard; peptide; 14 AA.
 XX ABB55905;
 XX 15-FEB-2002 (first entry)
 XX Vascular dementia-associated protein isoform (VPI) 105.
 XX Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
 XX diagnosis; prognosis; gene therapy.
 XX Homo sapiens.
 XX WO200169261-A2.
 XX 20-SEP-2001.
 XX 14-MAR-2001; 2001WO-GB001106.
 XX 15-MAR-2000; 2000GB-00006285.
 XX 24-NOV-2000; 2000GB-00028734.
 XX 28-NOV-2000; 2000US-00724391.
 XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
 XX Herath HWAC, Parekh RB, Rohlf C;
 XX WPI; 2001-557937/62.
 XX Screening, diagnosis or prognosis of vascular dementia (VD), useful for
 XX determining stage of VD and monitoring the effect of VD therapy,
 XX comprises analyzing body fluid by 2-dimensional electrophoresis for
 XX features correlated with VD.
 XX Claim 6; Page 32; 151pp; English.
 XX The invention relates to screening, diagnosis or prognosis of Vascular
 CC Dementia (VD) in a subject comprising analysing body fluid from the
 CC subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
 CC features containing at least one chosen feature whose relative abundance
 CC correlates with the presence, absence, stage or severity of VD or

CC predicts the onset or course of VD, especially detecting in a sample of
 CC cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
 CC protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
 CC specification. Detecting VD-associated features and VPI is useful for the
 CC screening, diagnosis or prognosis of VD, for determining the stage or
 CC severity of VD, for identifying a subject at risk of VD or for monitoring
 CC the effect of therapy administered to a subject having VD. Nucleic acids
 CC encoding a VPI or inhibiting the function of a VPI are useful for the
 CC treatment of VD and for gene therapy
 XX Sequence 14 AA;
 SQ

Query Match 2.0%; Score 14; DB 4; Length 14;
 Best Local Similarity 100.0%; Pred. No. 6.5e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 619 GSIQVDCBELVSGR 632
 DB 1 GSIQVDCBELVSGR 14

RESULT 8
 ABB55910
 ID ABB55910 standard; peptide; 10 AA.
 XX ABB55910;
 XX 15-FEB-2002 (first entry)
 XX Vascular dementia-associated protein isoform (VPI) 110.
 XX Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
 XX diagnosis; prognosis; gene therapy.
 XX Homo sapiens.
 XX WO200169261-A2.
 XX 20-SEP-2001.
 XX 14-MAR-2001; 2001WO-GB001106.
 XX 15-MAR-2000; 2000GB-00006285.
 XX 24-NOV-2000; 2000GB-00028734.
 XX 28-NOV-2000; 2000US-00724391.
 XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
 XX Herath HWAC, Parekh RB, Rohlf C;
 XX WPI; 2001-557937/62.
 XX Screening, diagnosis or prognosis of vascular dementia (VD), useful for
 XX determining stage of VD and monitoring the effect of VD therapy,
 XX comprises analyzing body fluid by 2-dimensional electrophoresis for
 XX features correlated with VD.
 XX Claim 6; Page 32; 151pp; English.
 XX The invention relates to screening, diagnosis or prognosis of Vascular
 CC Dementia (VD) in a subject comprising analysing body fluid from the
 CC subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
 CC features containing at least one chosen feature whose relative abundance
 CC correlates with the presence, absence, stage or severity of VD or
 CC predicts the onset or course of VD, especially detecting in a sample of
 CC cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
 CC protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
 CC specification. Detecting VD-associated features and VPI is useful for the
 CC screening, diagnosis or prognosis of VD, for determining the stage or
 CC severity of VD, for identifying a subject at risk of VD or for monitoring
 CC the effect of therapy administered to a subject having VD. Nucleic acids
 CC encoding a VPI or inhibiting the function of a VPI are useful for the
 CC treatment of VD and for gene therapy

XX SQ Sequence 10 AA;
 Query Match 1.4%; Score 10; DB 4; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.041;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 662 FSSGITGCVK 671
 |||||
 1 FSSGITGCVK 10

Db

RESULT 9
 ABB55909
 ID ABB55909 standard; peptide; 15 AA.
 AC ABB55909;
 XX
 DT 15-FEB-2002 (first entry)
 XX
 DE Vascular dementia-associated protein isoform (VPI) 109.
 XX
 KW Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
 KW diagnosis; prognosis; gene therapy.
 XX
 OS Homo sapiens.
 XX
 FN WO200169261-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 14-MAR-2001; 2001WO-GB001106.
 XX
 PR 15-MAR-2000; 2000GB-00006285.
 PR 24-NOV-2000; 2000GB-00028734.
 PR 28-NOV-2000; 2000US-00724391.
 XX
 PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
 XX
 PI Herath HMAC, Parekh RB, Rohlf C;
 DR WPI; 2001-557937/62.
 XX
 PT Screening, diagnosis or prognosis of vascular dementia (VD), useful for
 PT determining stage of VD and monitoring the effect of VD therapy, for
 PT comprises analyzing body fluid by 2-dimensional electrophoresis, for
 PT features correlated with VD.
 XX
 PS Claim 6; Page 32; 151pp; English.
 XX
 CC The invention relates to screening, diagnosis or prognosis of Vascular
 CC Dementia (VD) in a subject comprising analysing body fluid from the
 CC subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
 CC features containing at least one chosen feature whose relative abundance
 CC correlates with the presence, absence, stage or severity of VD or
 CC predicts the onset or course of VD, especially detecting in a sample of
 CC cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
 CC protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
 CC specification. Detecting VD-associated features and VPI is useful for the
 CC screening, diagnosis or prognosis of VD, for determining the stage or
 CC severity of VD, for identifying a subject at risk of VD or for monitoring
 CC the effect of therapy administered to a subject having VD. Nucleic acids
 CC encoding a VPI or inhibiting the function of a VPI are useful for the
 CC treatment of VD and for gene therapy
 XX
 SQ Sequence 15 AA;
 Query Match 1.3%; Score 9; DB 4; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.58;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 587 YQLSGGEAR 595
 |||||

Db 1 YQLSGGEAR 9

RESULT 10
 ABM65732
 ID ABM65732 standard; protein; 159 AA.
 XX
 AC ABM65732;
 XX
 DT 20-OCT-2003 (first entry)
 XX
 DE Propionibacterium acnes immunogenic polypeptide #30408.
 XX
 KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;
 KW immunoestimulant; immune response; vaccine; immunogenic.
 XX
 OS Propionibacterium acnes.
 XX
 FN WO2003033515-A1.
 XX
 PD 24-APR-2003.
 XX
 PF 11-OCT-2002; 2002WO-US032727.
 XX
 PR 15-OCT-2001; 2001US-00978825.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
 PI Barth B, Vallieve-Douglas J;
 XX
 DR WPI; 2003-381789/36.
 XX
 PT New Propionibacterium acnes polypeptides and polynucleotides encoding the
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
 PT or for stimulating an immune response specific for a P. acnes protein.
 XX
 PS Claim 7; SEQ ID NO 30408; 1481pp; English.
 XX
 CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
 CC encoding a Propionibacterium acnes protein. The invention also relates to
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
 CC immunogenic fragments of P. acnes polypeptides. The invention
 CC additionally encompasses expression vectors and host cells comprising a
 CC polynucleotide of the invention; antibodies against polypeptides of the
 CC invention; fusion proteins comprising a polypeptide of the invention; a
 CC method for stimulating an immune response specific for a P. acnes
 CC polypeptide and an isolated T cell population comprising T cells prepared
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or
 CC antigen-presenting cells that express the polypeptide); a method and kit
 CC for detecting or determining the presence or absence of P. acnes in a
 CC patient; and a method for inhibiting the development of P. acnes in a
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
 CC proteins, T cell populations or antigen-presenting cells that express the
 CC polypeptides are useful for diagnosing, preventing or treating acne
 CC vulgaris, or for stimulating an immune response specific for a P. acnes
 CC protein. The polynucleotides can also be used as probes or primers for
 CC nucleic acid hybridisation. The vaccine composition is useful for the
 CC stimulation of an immune response against P. acnes, or for treating acne,
 CC and the kit is useful for performing a diagnostic assay. The present
 CC sequence represents a specifically claimed P. acnes polypeptide which is
 CC thought to contain an immunogenic region. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 159 AA;
 Query Match 1.3%; Score 9; DB 6; Length 159;
 Best Local Similarity 100.0%; Pred. No. 5.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 548 VRTSTASGL 556
 DB 31 VRTSTASGL 39

Query Match 1.3%; Score 9; DB 7; Length 2531;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGTCQ 474
 DB 1029 PCLHGTCQ 1037

RESULT 12
 ADE63705
 ID ADE63705 standard; protein; 2531 AA.
 XX
 AC ADE63705;
 DT 29-JAN-2004 (first entry)
 DE Rat Protein CAA40667, SEQ ID NO 9649.
 KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 OS Rattus norvegicus.
 PN WO2003016475-A2.
 PD 27-FEB-2003.
 PF 14-AUG-2002; 2002WO-US025765.
 PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 PR (GEO) GEN HOSPITAL CORP.
 PA (PAB) BAYER AG.
 PI Woolf C, D'urso D, Befort K, Costigan M;
 XX WPI; 2003-268312/26.
 DR GENBANK; CAA40667.
 XX

New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.

Claim 1; Page; 1017pp; English.

The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a rat protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: ftp.wipo.int/pub/published_pct_sequences.

Sequence 2531 AA;

CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2531 AA;
Query Match 1.3%; Score 9; DB 7; Length 2531;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 466 PCLHGTCQ 474
Db 1029 PCLHGTCQ 1037
RESULT 13
ADE63709
ID ADE63709 standard; protein; 2531 AA.
AC ADE63709;
XX
XX
DT 29-JAN-2004 (first entry)
XX
XX Rat Protein CAA40667, SEQ ID NO 9653.
DE
XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
XX chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
XX Rattus norvegicus.
XX
XX WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
XX 14-AUG-2002; 2002WO-US025765.
XX
XX 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
XX (GEO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
PA
XX
XX Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
DR GENBANK; CAA40667.
XX
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
XX Claim 1; Page; 1017pp; English.
XX
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition,
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction

CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2531 AA;
Query Match 1.3%; Score 9; DB 7; Length 2531;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 466 PCLHGTCQ 474
Db 1029 PCLHGTCQ 1037
RESULT 14
ADE63701
ID ADE63701 standard; protein; 2531 AA.
XX
XX ADE63701;
XX
XX 29-JAN-2004 (first entry)
XX
XX Rat Protein CAA40667, SEQ ID NO 9645.
DE
XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
XX chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
XX Rattus norvegicus.
XX
XX WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
XX 14-AUG-2002; 2002WO-US025765.
XX
XX 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
XX (GEO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
PA
XX
XX Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
DR GENBANK; CAA40667.
XX
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
XX Claim 1; Page; 1017pp; English.
XX
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition,
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction

CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2531 AA;

Query Match 1.3%; Score 9; DB 7; Length 2531;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGGTCQ 474
 |||||
 DB 1029 PCLHGGTCQ 1037

RESULT 15

ABB70376
 ID ABB70376 standard; protein; 3319 AA.

XX
 AC ABB70376;

XX
 DT 26-MAR-2002 (first entry)

XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 37920.

XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX
 OS Drosophila melanogaster.

XX
 PN WO200171042-A2.

XX
 PD 27-SEP-2001.

XX
 PF 23-MAR-2001; 2001WO-US009231.

XX
 PR 23-MAR-2000; 2000US-0191637P.

XX
 PR 11-JUL-2000; 2000US-00614150.

XX
 PA (PEKE) PE CORP NY.

XX
 FI Venter JC, Adams M, Li PWD, Myers EW;

XX
 DR WPI; 2001-656960/75.

XX
 DR N-PSDB; ABL14479.

XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.

XX
 PS Disclosure; SEQ ID NO 37920; 21pp + Sequence Listing; English.

XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABU16176-ABU16175), expressed DNA
 CC sequences (ABU161840-ABU16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 3319 AA;

Query Match 1.3%; Score 9; DB 4; Length 3319;

Best Local Similarity 100.0%; Pred. No. 96;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 603 NDGEWHRV 611

|||||

DB 3033 NDGEWHRV 3041

Search completed: March 9, 2004, 17:27:53
 Job time : 88.8465 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:25:24 ; Search time 27.0475 Seconds
(without alignments)
1345.642 Million cell updates/sec

Title: US-10-006-011A-3
Perfect score: 705
Sequence: 1 EKITFRPDSADGMLYNGQ.....QPLDLQHRQAQAGANTPCPS 705

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/PTUS COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfilesi.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.3	1940	2	US-08-644-271-30
2	9	1.3	1940	4	US-09-077-955-34
3	8	1.1	241	4	US-09-107-532A-5434
4	8	1.1	427	4	US-09-232-991A-19300
5	8	1.1	579	4	US-09-540-236-2071
6	8	1.1	1015	3	US-08-537-210A-1
7	8	1.1	1015	3	US-09-113-825-1
8	8	1.1	1130	2	US-08-460-309-2
9	8	1.1	1130	2	US-08-125-077-2
10	8	1.1	1130	6	5444158-2
11	8	1.1	1251	1	US-08-340-428B-49
12	8	1.1	2321	4	US-09-230-652-2
13	8	1.1	2471	1	US-08-185-432-16
14	8	1.1	2471	1	US-08-083-590A-19
15	8	1.1	2471	3	US-08-532-384-19
16	8	1.1	2471	4	US-08-899-232-1
17	8	1.1	3084	4	US-09-562-702A-12
18	8	1.1	3088	4	US-09-562-702A-8
19	8	1.1	3089	4	US-09-562-702A-4
20	8	1.1	3106	4	US-09-562-702A-10
21	8	1.1	3110	4	US-09-562-702A-2
22	8	1.1	3110	4	US-09-562-702A-6
23	8	1.1	3110	4	US-09-561-709B-7
24	8	1.1	3111	2	US-08-460-309-4
25	8	1.1	3111	2	US-08-125-077-4
26	7	1.0	77	4	US-09-252-991A-18061
27	7	1.0	100	4	US-09-249-697A-3

28	7	1.0	100	4	US-09-363-316B-3
29	7	1.0	108	4	US-09-621-976-4170
30	7	1.0	146	4	US-09-543-681A-7317
31	7	1.0	162	4	US-09-107-532A-5488
32	7	1.0	165	4	US-09-252-991A-27759
33	7	1.0	181	4	US-09-252-991A-23483
34	7	1.0	186	4	US-09-252-991A-21008
35	7	1.0	186	4	US-09-107-532A-6672
36	7	1.0	201	4	US-09-252-991A-28786
37	7	1.0	216	1	US-08-315-695-20
38	7	1.0	237	4	US-09-252-991A-28116
39	7	1.0	240	4	US-09-570-856B-16
40	7	1.0	244	4	US-09-252-991A-18465
41	7	1.0	248	2	US-08-755-559-1
42	7	1.0	248	3	US-09-210-474-1
43	7	1.0	248	4	US-09-539-774-1
44	7	1.0	273	4	US-09-252-991A-28324
45	7	1.0	290	4	US-09-252-991A-27099

ALIGNMENTS

RESULT 1
US-08-644-271-30
; Sequence 30, Application US/08644271
; Patent No. 5814478
; GENERAL INFORMATION:
; APPLICANT: Valenzuela, et al.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
; TITLE OF INVENTION: AND LIGANDS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/644,271
FILING DATE: 10-MAY-1996
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: USSN 60/008,657
FILING DATE: 15-DEC-1995

ATTORNEY/AGENT INFORMATION:
NAME: Cobert, Robert J
REGISTRATION NUMBER: 36,108

REFERENCE/DOCKET NUMBER: REG 195A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-345-7400
TELEFAX: 914-345-7721

TELEX:
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 1940 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Rat Agrin
LOCATION: 1...1940
OTHER INFORMATION:
US-08-644-271-30

Query Match 1.3%; Score 9; DB 2; Length 1940;

Best Local Similarity 100.0%; Pred. No. 20;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGTCQ 474
Db 1228 PCLHGTCQ 1236

RESULT 2
US-09-077-955-34
; Sequence 34, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; EARLIER FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
; LENGTH: 1940
; TYPE: PRT
; ORGANISM: Rattus sp.
US-09-077-955-34

Query Match 1.3%; Score 9; DB 4; Length 1940;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGTCQ 474
Db 1228 PCLHGTCQ 1236

RESULT 3
US-09-107-532A-5434
; Sequence 5434, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneka
; REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5434:
SEQUENCE CHARACTERISTICS:
LENGTH: 241 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...241
SEQUENCE DESCRIPTION: SEQ ID NO: 5434:
US-09-107-532A-5434

Query Match 1.1%; Score 8; DB 4; Length 241;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 IVFHDNL 151
Db 28 IVFHDNL 35

RESULT 4
US-09-252-991A-19300
; Sequence 19300, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19300
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19300

Query Match 1.1%; Score 8; DB 4; Length 427;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 LLSLTQ 83
Db 247 LLSLTQ 254

RESULT 5
US-09-540-236-2071
; Sequence 2071, Application US/09540236
; Patent No. 6873910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2071
; LENGTH: 579


```
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-2071

Query Match          1.1%; Score 8; DB 4; Length 579;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      319 AEPLALGR 326
      |||||
Db      320 AEPLALGR 327

RESULT 6
US-08-537-210A-1
; Sequence 1, Application US/08537210A
; Patent No. 5780300
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Fortini, Mark
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
; TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/537,210A
; FILING DATE: 29-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1015 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: hum N (Human No. 5780300ch 2)
; LOCATION: 1155...2169
; OTHER INFORMATION: Highly conserved ankyrin repeat
; OTHER INFORMATION: region of No. 5780300ch
US-08-537-210A-1

Query Match          1.1%; Score 8; DB 1; Length 1015;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      426 PCQHGATC 433
      |||||
Db      3 PCQHGATC 10

RESULT 7
US-09-113-825-1
; Sequence 1, Application US/09113825
; Patent No. 6149902
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Fortini, Mark
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
; TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/113,825
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/537,210
; FILING DATE: 29-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1015 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: hum N (Human No. 6149902ch 2)
; LOCATION: 1155...2169
; OTHER INFORMATION: Highly conserved ankyrin repeat
; OTHER INFORMATION: region of No. 6149902ch
US-09-113-825-1

Query Match          1.1%; Score 8; DB 3; Length 1015;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      426 PCQHGATC 433
      |||||
Db      3 PCQHGATC 10

RESULT 8
US-08-460-309-2
; Sequence 2, Application US/08460309
; Patent No. 5837496
; GENERAL INFORMATION:
; APPLICANT: Engvall, Eva
; APPLICANT: Leivo, Ilmo
; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
; TITLE OF INVENTION: Fragments and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
```

```

; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,309
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/125,077
; FILING DATE: 22-SEP-1993
; APPLICATION NUMBER: US PCT/US 94/10730
; FILING DATE: 21-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/472,319
; FILING DATE: 30-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/919,951
; FILING DATE: 27-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 9721
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-8949
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1130 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-460-309-2

Query Match 1.1%; Score 8; DB 2; Length 1130;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 543 TIELEVRT 550
DB 808 TIELEVRT 815

RESULT 9
US-08-125-077-2
; Sequence 2, Application US/08125077
; Patent No. 5872231
; Patent No. 5872231 5840863
; GENERAL INFORMATION:
; APPLICANT: Engvall, Eva
; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
; TITLE OF INVENTION: Fragments and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/125,077
; FILING DATE: 22-SEP-1993

```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US 94/10730
; FILING DATE: 21-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/472,319
; FILING DATE: 30-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/919,951
; FILING DATE: 27-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 9721
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1130 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; US-08-125-077-2

Query Match 1.1%; Score 8; DB 2; Length 1130;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 543 TIELEVRT 550
DB 808 TIELEVRT 815

RESULT 10
5444158-2
; Patent No. 5444158
; APPLICANT: ENGVALL, EVA; SANES, JOSHUA
; TITLE OF INVENTION: MEROSIN, NUCLEIC ACIDS ENCODING,
; FRAGMENTS AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/97,642
; FILING DATE: 08-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 587,689
; FILING DATE: 24-SEP-1990
; APPLICATION NUMBER: 472,319
; FILING DATE: 30-JAN-1990
; SEQ ID NO: 2:
; LENGTH: 1130
5444158-2

Query Match 1.1%; Score 8; DB 6; Length 1130;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 543 TIELEVRT 550
DB 808 TIELEVRT 815

RESULT 11
US-08-340-428B-49
; Sequence 49, Application US/08340428B
; Patent No. 5648465
; GENERAL INFORMATION:
; APPLICANT: MARGOLIS, Richard U.
; APPLICANT: RAUCH, Uwe
; APPLICANT: MARGOLIS, Renee K.
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES FOR A
; TITLE OF INVENTION: NEUROCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN
; NUMBER OF SEQUENCES: 49

```

/ CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Browdy and Neimark
 / STREET: 419 Seventh Street, N.W.
 / CITY: Washington
 / STATE: D.C.
 / COUNTRY: U.S.A.
 / ZIP: 20004
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: Patent In Release #1.0, Version #1.25
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/340,428B
 / FILING DATE: 14 No. 5648465ember 1994
 / CLASSIFICATION: 514
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: 07/922,911
 / FILING DATE: 03 August 1992
 / CLASSIFICATION: 514
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Browdy, Roger L.
 / REGISTRATION/DOCKET NUMBER: 25,618
 / REFERENCE/DOCKET NUMBER: Margolis=1A
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: 202-628-5197
 / TELEFAX: 202-737-3528
 / INFORMATION FOR SEQ ID NO: 49:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 1257 amino acids
 / TYPE: amino acid
 / STRANDEDNESS: single
 / TOPOLOGY: linear
 / MOLECULE TYPE: peptide
 / US-08-340-428B-49

```
Query Match      1.1%; Score 8; DB 1; Length 1257;
Best Local Similarity 100.0%; Pred.No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels
```

Qy 466 PCLHGGTC 473
D_b 957 PCLHGGTC 964

RESULT 12
 US-09-230-652-2
 ; Sequence 2, Application US/09230652A
 ; Patent No. 6537775
 ; GENERAL INFORMATION:
 ; APPLICANT: Tournier-Lasserre, Elisabeth
 ; APPLICANT: Joutel, Anne
 ; APPLICANT: Bousser, Marie-Germaine
 ; APPLICANT: Bach, Jean-Francois
 ; TITLE OF INVENTION: GENE INVOLVED IN CADASIL,
 ; METHOD OF DIAGNOSIS AND
 ; TITLE OF INVENTION: THERAPEUTIC APPLICATION
 ; FILE REFERENCE: 03715.0048-00000
 ; CURRENT APPLICATION NUMBER: US/09/230,652A
 ; CURRENT FILING DATE: 1999-05-17
 ; EARLIER APPLICATION NUMBER: FR 96 09733
 ; EARLIER FILING DATE: 1996-08-01
 ; EARLIER APPLICATION NUMBER: FR 97 04680
 ; EARLIER FILING DATE: 1997-04-16
 ; EARLIER APPLICATION NUMBER: PCT/FR97/01433
 ; EARLIER FILING DATE: 1997-07-31
 ; NUMBER OF SEQ ID NOS: 163
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 2321
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: human ADNC No. 6537775sch 3

US-09-230-652-2

```
Query Match      1.1%; Score 8; DB 4; Length 2321;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 483 GFSGPRCQ 490
Db 1237 GFSGPRCQ 1244

RESULT 13
US-08-185-432-16
Sequence 16, Application US/08185432
Patent No. 5750852
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Busseau, Isabelle
APPLICANT: Diederich, Robert J.
APPLICANT: Xu, Tian
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: DELTEx PROTEINS, NUCLEIC ACIDS, AND
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatenIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2471 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-16

```
Query Match      1.1%; Score 8; DB 1; Length 2471;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 426 PCQHGATC 433
db 1157 PCQHGATC 1164

RESULT 14
US-08-083-590A-19
; Sequence 19, Application US/08083590A
; Patent No. 5796159
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S.
; TITLE OF INVENTION: Therapeutic Agent
; TITLE OF INVENTION: And Composition

```
; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2471 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-083-590A-19

Query Match 1.1%; Score 8; DB 1; Length 2471;
Best Local Similarity 100.0%; Pred.No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 426 PCQHGATC 433
Db 1157 PCQHGATC 1164

Search completed: March 9, 2004, 17:32:24
Job time : 28.0475 secs

; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2471 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-532-384-19

Query Match 1.1%; Score 8; DB 3; Length 2471;
Best Local Similarity 100.0%; Pred.No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 426 PCQHGATC 433
Db 1157 PCQHGATC 1164

Search completed: March 9, 2004, 17:32:24
Job time : 28.0475 secs

RESULT 15
US-08-532-384-19
; Sequence 19, Application US/08532384
; Patent No. 6083904
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/532,384
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/083,590
; FILING DATE: 25-JUN-1993
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:27:59 ; Search time 51.7431 Seconds
(without alignments)
2876.963 Million cell updates/sec

Title: US-10-006-011A-3
Perfect score: 705
Sequence: 1 EKIIFRPSADGMLLYNCQ.....QPLDLQRAQAGNTRPCPS 705

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 809742 seqs, 211153259 residues

Word size : 0
Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/prodata1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	1.3	1940	13	US-10-016-283-34
2	9	1.3	2531	15	US-10-190-115-29
3	9	1.3	2531	15	US-10-369-072-29
4	8	1.1	54	10	US-09-776-724A-136
5	8	1.1	80	9	US-09-864-761-35784
6	8	1.1	131	15	US-10-264-237-2547
7	8	1.1	264	11	US-09-833-245-2022
8	8	1.1	264	15	US-10-266-829-75
9	8	1.1	299	14	US-10-219-220-251
10	8	1.1	309	14	US-10-122-706-31
11	8	1.1	383	11	US-09-833-245-770
12	8	1.1	383	14	US-10-245-103-80
13	8	1.1	383	14	US-10-245-107-80
14	8	1.1	383	14	US-10-245-143-80
15	8	1.1	383	14	US-10-245-771-80

16	8	1.1	383	14	US-10-245-851-80	Sequence 80, Appl
17	8	1.1	383	14	US-10-245-883-80	Sequence 80, Appl
18	8	1.1	383	14	US-10-237-535-80	Sequence 80, Appl
19	8	1.1	383	14	US-10-238-183-80	Sequence 80, Appl
20	8	1.1	383	14	US-10-238-283-80	Sequence 80, Appl
21	8	1.1	383	14	US-10-238-370-80	Sequence 80, Appl
22	8	1.1	383	14	US-10-245-055-80	Sequence 80, Appl
23	8	1.1	383	14	US-10-245-147-80	Sequence 80, Appl
24	8	1.1	383	14	US-10-245-730-80	Sequence 80, Appl
25	8	1.1	383	14	US-10-245-739-80	Sequence 80, Appl
26	8	1.1	383	14	US-10-246-210-80	Sequence 80, Appl
27	8	1.1	383	14	US-10-239-196-80	Sequence 80, Appl
28	8	1.1	383	14	US-10-243-024-80	Sequence 80, Appl
29	8	1.1	383	14	US-10-243-409-80	Sequence 80, Appl
30	8	1.1	383	14	US-10-245-621-80	Sequence 80, Appl
31	8	1.1	383	14	US-10-245-880-80	Sequence 80, Appl
32	8	1.1	383	14	US-10-245-033-80	Sequence 80, Appl
33	8	1.1	383	14	US-10-243-095-80	Sequence 80, Appl
34	8	1.1	383	14	US-10-245-185-80	Sequence 80, Appl
35	8	1.1	383	14	US-10-245-427-80	Sequence 80, Appl
36	8	1.1	383	14	US-10-245-473-80	Sequence 80, Appl
37	8	1.1	383	14	US-10-245-770-80	Sequence 80, Appl
38	8	1.1	383	14	US-10-245-877-80	Sequence 80, Appl
39	8	1.1	383	14	US-10-246-976-80	Sequence 80, Appl
40	8	1.1	383	14	US-10-243-320-80	Sequence 80, Appl
41	8	1.1	383	14	US-10-242-743-80	Sequence 80, Appl
42	8	1.1	383	14	US-10-242-845-80	Sequence 80, Appl
43	8	1.1	383	14	US-10-237-636-80	Sequence 80, Appl
44	8	1.1	383	14	US-10-238-325-80	Sequence 80, Appl
45	8	1.1	383	14	US-10-238-346-80	Sequence 80, Appl

ALIGNMENTS

RESULT 1
US-10-016-283-34
; Sequence 34, Application US/10016283
; Publication No. US20020164702A1
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/10/016,283
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US/09/077,955A
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: PCT/US96/20696
; PRIOR FILING DATE: 1996-12-13
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
; LENGTH: 1940
; TYPE: PRT
; ORGANISM: Rattus sp.
US-10-016-283-34

Query Match	1.3%	Score 9;	DB 13;	Length 1940;
Best Local Similarity	100.0%	Pred. No. 45;		
Matches	9;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	466	PCLHGGTCQ 474		
Db	1228	PCLHGGTCQ 1236		
RESULT 2				
US-10-190-115-29				
; Sequence 29, Application US/10190115				
; Publication No. US20030207394A1				
; GENERAL INFORMATION:				
; APPLICANT: Alsobrook, John P. II				
; APPLICANT: Boldog, Ferenc L.				

; APPLICANT: Burgess, Catherine E.
 ; APPLICANT: Caeman, Stacie J.
 ; APPLICANT: Grosse, William M.
 ; APPLICANT: Gusev, Vladimir Y.
 ; APPLICANT: Ji, Weizhen
 ; APPLICANT: Lepley, Denise M.
 ; APPLICANT: Liu, Xiaohong
 ; APPLICANT: Mezick, Amanda J.
 ; APPLICANT: Padigaru, Muralidhara
 ; APPLICANT: Rastelli, Luca
 ; APPLICANT: Shen, Lei
 ; APPLICANT: Shenoy, Suresh G.
 ; APPLICANT: Shimkets, Richard A.
 ; APPLICANT: Spaderna, Steven K.
 ; APPLICANT: Spytek, Kimberly A.
 ; APPLICANT: Szerkes, Edward S. Jr.
 ; APPLICANT: Taupier, Raymond J. Jr.
 ; APPLICANT: Tchernev, Velizar T.
 ; APPLICANT: Zerhusen, Bryan D.
 ; APPLICANT: Voss, Edward Z.
 ; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
 ; FILE REFERENCE: 21402-050 CIP
 ; CURRENT APPLICATION NUMBER: US/10/190,115
 ; CURRENT FILING DATE: 2003-02-10
 ; PRIOR APPLICATION NUMBER: 60/303,168
 ; PRIOR FILING DATE: 2001-07-05
 ; PRIOR APPLICATION NUMBER: 60/368,996
 ; PRIOR FILING DATE: 2002-04-01
 ; PRIOR APPLICATION NUMBER: 60/386,816
 ; PRIOR FILING DATE: 2002-06-07
 ; PRIOR APPLICATION NUMBER: 60/215,854
 ; PRIOR FILING DATE: 2000-07-03
 ; PRIOR APPLICATION NUMBER: 60/215,856
 ; PRIOR FILING DATE: 2000-07-03
 ; PRIOR APPLICATION NUMBER: 60/215,902
 ; PRIOR FILING DATE: 2000-07-03
 ; PRIOR APPLICATION NUMBER: 60/216,585
 ; PRIOR FILING DATE: 2000-07-07
 ; PRIOR APPLICATION NUMBER: 60/386,816
 ; PRIOR FILING DATE: 2002-04-01
 ; PRIOR APPLICATION NUMBER: 60/215,854
 ; PRIOR FILING DATE: 2000-07-03
 ; PRIOR APPLICATION NUMBER: 60/215,856
 ; PRIOR FILING DATE: 2000-07-03
 ; PRIOR APPLICATION NUMBER: 60/215,902
 ; PRIOR FILING DATE: 2000-07-03
 ; PRIOR APPLICATION NUMBER: 60/216,585
 ; PRIOR FILING DATE: 2001-07-07
 ; PRIOR APPLICATION NUMBER: 60/216,722
 ; PRIOR FILING DATE: 2000-07-07
 ; PRIOR APPLICATION NUMBER: 60/218,622
 ; PRIOR FILING DATE: 2000-07-17
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 136
 ; SOFTWARE: CuraSeqList version 0.1
 ; SEQ ID NO 29
 ; LENGTH: 2531
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-190-115-29

Query Match 1.3%; Score 9; DB 15; Length 2531;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGGTCQ 474
 DB 1029 PCLHGGTCQ 1037

RESULT 3
 US-10-369-072-29
 ; Sequence 29, Application US/10369072
 ; Publication No. US20040014081A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Alsobrook II, John P
 ; APPLICANT: Spaderna, Stephen K
 ; APPLICANT: Tchernev, Velizar
 ; APPLICANT: Liu, Xiaohong
 ; APPLICANT: Shenoy, Suresh

; APPLICANT: Spytek, Kimberly
 ; APPLICANT: Zerhusen, Bryan
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Taupier, Raymond T
 ; APPLICANT: Rastelli, Luca
 ; APPLICANT: Grosse, William M
 ; APPLICANT: Szerkes, Edward S
 ; APPLICANT: Lepley, Denise M
 ; APPLICANT: Shen, Lei
 ; APPLICANT: Burgess, Catherine E
 ; APPLICANT: Shimkets, Richard
 ; APPLICANT: Padigaru, Muralidhara
 ; TITLE OF INVENTION: No. US20040014081A1el Proteins and Nucleic Acids Encoding Same
 ; FILE REFERENCE: 21402-050 CON2
 ; CURRENT APPLICATION NUMBER: US/10/369,072
 ; CURRENT FILING DATE: 2003-02-18
 ; PRIOR APPLICATION NUMBER: 10/174,372
 ; PRIOR FILING DATE: 2002-06-17
 ; PRIOR APPLICATION NUMBER: 09/898,994
 ; PRIOR FILING DATE: 2001-07-03
 ; PRIOR APPLICATION NUMBER: 60/215,854
 ; PRIOR FILING DATE: 2000-07-03
 ; PRIOR APPLICATION NUMBER: 60/215,856
 ; PRIOR FILING DATE: 2000-07-03
 ; PRIOR APPLICATION NUMBER: 60/215,902
 ; PRIOR FILING DATE: 2000-07-03
 ; PRIOR APPLICATION NUMBER: 60/216,585
 ; PRIOR FILING DATE: 2000-07-07
 ; PRIOR APPLICATION NUMBER: 60/216,586
 ; PRIOR FILING DATE: 2000-07-07
 ; PRIOR APPLICATION NUMBER: 60/216,722
 ; PRIOR FILING DATE: 2000-07-07
 ; PRIOR APPLICATION NUMBER: 60/218,622
 ; PRIOR FILING DATE: 2000-07-17
 ; PRIOR APPLICATION NUMBER: 60/218,992
 ; PRIOR FILING DATE: 2000-07-17
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 100
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 29
 ; LENGTH: 2531
 ; TYPE: PRT
 ; ORGANISM: Rattus norvegicus
 US-10-369-072-29

Query Match 1.3%; Score 9; DB 15; Length 2531;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGGTCQ 474
 DB 1029 PCLHGGTCQ 1037

RESULT 4
 US-09-776-724A-136
 ; Sequence 136, Application US/09776724A
 ; Publication No. US20030050455A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: 64 Human Secreted Proteins
 ; FILE REFERENCE: P2011
 ; CURRENT APPLICATION NUMBER: US/09/776,724A
 ; CURRENT FILING DATE: 2001-02-06
 ; PRIOR APPLICATION NUMBER: 60/180,909
 ; PRIOR FILING DATE: 2000-02-08
 ; PRIOR APPLICATION NUMBER: 09/669,688
 ; PRIOR FILING DATE: 2000-09-26
 ; PRIOR APPLICATION NUMBER: 09/229,982
 ; PRIOR FILING DATE: 1999-01-14
 ; PRIOR APPLICATION NUMBER: PCT/US98/14613
 ; PRIOR FILING DATE: 1998-07-15
 ; PRIOR APPLICATION NUMBER: 60/052,661

```

; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: 60/052,872
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: 60/052,871
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: 60/052,874
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: 60/052,873
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: 60/052,870
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: 60/052,875
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: 60/053,440
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,441
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,442
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/056,359
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,725
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,985
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,952
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,989
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/056,361
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,726
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,724
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,946
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,683
; PRIOR FILING DATE: 1997-08-18
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 136
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (32)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (39)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-776-724A-136

Query Match 1.1%; Score 8; DB 10; Length 54;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 74 VTLRLSLT 81
Db 47 VTLRLSLT 54

RESULT 5
US-09-864-761-35784
; Sequence 35784, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Harzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

```

```

; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35784
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL035413.19
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3
; OTHER INFORMATION: EST_HUMAN HIT: BE875511.1, EVALUATE 3.00e-14
US-09-864-761-35784

Query Match 1.1%; Score 8; DB 9; Length 80;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 680 PGAPPPQP 687
Db 33 PGAPPPQP 40

RESULT 6
US-10-264-237-2547
; Sequence 2547, Application US/10264237

```


Publication No. US20040009491A1
 GENERAL INFORMATION:
 APPLICANT: Birse et al.
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 FILE REFERENCE: P4131P1
 CURRENT APPLICATION NUMBER: US/10/264,237
 CURRENT FILING DATE: 2002-10-04
 PRIOR APPLICATION NUMBER: PCT/US01/16450
 PRIOR FILING DATE: 2001-05-18
 PRIOR APPLICATION NUMBER: US 60/205,515
 PRIOR FILING DATE: 2000-05-19
 NUMBER OF SEQ ID NOS: 2876
 SOFTWARE: PatentIn Ver. 3.1
 SEQ ID NO 2547
 LENGTH: 131
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: MISC FEATURE
 LOCATION: (106)
 OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
 FEATURE:
 NAME/KEY: MISC FEATURE
 LOCATION: (108)
 OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
 FEATURE:
 NAME/KEY: MISC FEATURE
 LOCATION: (115)
 OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
 FEATURE:
 NAME/KEY: MISC FEATURE
 LOCATION: (120)
 OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
 FEATURE:
 NAME/KEY: MISC FEATURE
 LOCATION: (127)
 OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
 US-10-264-237-2547

Query Match 1.1%; Score 8; DB 15; Length 131;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 GVLFFSGG 285
 DB 58 GVLFFSGG 65

RESULT 7
 US-09-833-245-2022
 Sequence 2022, Application US/09833245
 Publication No. US20040010134A1
 GENERAL INFORMATION:
 APPLICANT: Human Genome Sciences, Inc.
 TITLE OF INVENTION: Albumin Fusion Proteins
 FILE REFERENCE: PF546PCT
 CURRENT APPLICATION NUMBER: US/09/833,245
 CURRENT FILING DATE: 2001-04-12
 PRIOR APPLICATION NUMBER: 60/229, 358
 PRIOR FILING DATE: 2000-04-12
 PRIOR APPLICATION NUMBER: 60/256, 931
 PRIOR FILING DATE: 2000-12-21
 PRIOR APPLICATION NUMBER: 60/199, 384
 PRIOR FILING DATE: 2000-04-25
 NUMBER OF SEQ ID NOS: 2267
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 2022
 LENGTH: 264
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-833-245-2022

Query Match 1.1%; Score 8; DB 11; Length 264;

Best Local Similarity 100.0%; Pred. No. 66;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 GVLFFSGG 285
 DB 13 GVLFFSGG 20

RESULT 8
 US-10-266-829-75
 Sequence 75, Application US/10266829
 Publication No. US20030220489A1
 GENERAL INFORMATION:
 APPLICANT: Rosen et al.
 TITLE OF INVENTION: 29 Human secreted proteins
 FILE REFERENCE: P2041P1
 CURRENT APPLICATION NUMBER: US/10/266,829
 CURRENT FILING DATE: 2002-10-09
 PRIOR APPLICATION NUMBER: 09/756,168
 PRIOR FILING DATE: 2001-01-09
 PRIOR APPLICATION NUMBER: PCT/US00/19735
 PRIOR FILING DATE: 2000-07-20
 PRIOR APPLICATION NUMBER: 60/145,220
 PRIOR FILING DATE: 1999-07-23
 NUMBER OF SEQ ID NOS: 146
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 75
 LENGTH: 264
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-266-829-75

Query Match 1.1%; Score 8; DB 15; Length 264;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 GVLFFSGG 285
 DB 13 GVLFFSGG 20

RESULT 9
 US-10-219-220-251
 Sequence 251, Application US/10219220
 Publication No. US20030082724A1
 GENERAL INFORMATION:
 APPLICANT: Flinn, Barry
 TITLE OF INVENTION: Compositions affecting programmed cell
 death and their use in the modification of plant development
 FILE REFERENCE: 11000.1022c1
 CURRENT APPLICATION NUMBER: US/10/219,220
 CURRENT FILING DATE: 2002-08-14
 PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
 PRIOR FILING DATE: 1999-06-04
 NUMBER OF SEQ ID NOS: 290
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 251
 LENGTH: 299
 TYPE: PRT
 ORGANISM: Pinus radiata
 US-10-219-220-251

Query Match 1.1%; Score 8; DB 14; Length 299;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 VTLRLSLT 81
 DB 260 VTLRLSLT 267

RESULT 10

Publication No. US20040009491A1
 GENERAL INFORMATION:
 APPLICANT: Birse et al.
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 FILE REFERENCE: P4131P1
 CURRENT APPLICATION NUMBER: US/10/264,237
 CURRENT FILING DATE: 2002-10-04
 PRIOR APPLICATION NUMBER: PCT/US01/16450
 PRIOR FILING DATE: 2001-05-18
 PRIOR APPLICATION NUMBER: US 60/205,515
 PRIOR FILING DATE: 2000-05-19
 NUMBER OF SEQ ID NOS: 2876
 SOFTWARE: PatentIn Ver. 3.1
 SEQ ID NO 2547
 LENGTH: 131
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: MISC FEATURE
 LOCATION: (106)
 OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
 FEATURE:
 NAME/KEY: MISC FEATURE
 LOCATION: (108)
 OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
 FEATURE:
 NAME/KEY: MISC FEATURE
 LOCATION: (115)
 OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
 FEATURE:
 NAME/KEY: MISC FEATURE
 LOCATION: (120)
 OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
 FEATURE:
 NAME/KEY: MISC FEATURE
 LOCATION: (127)
 OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
 US-10-264-237-2547

Query Match 1.1%; Score 8; DB 15; Length 131;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 GVLFFSGG 285
 DB 58 GVLFFSGG 65

RESULT 7
 US-09-833-245-2022
 Sequence 2022, Application US/09833245
 Publication No. US20040010134A1
 GENERAL INFORMATION:
 APPLICANT: Human Genome Sciences, Inc.
 TITLE OF INVENTION: Albumin Fusion Proteins
 FILE REFERENCE: PF546PCT
 CURRENT APPLICATION NUMBER: US/09/833,245
 CURRENT FILING DATE: 2001-04-12
 PRIOR APPLICATION NUMBER: 60/229, 358
 PRIOR FILING DATE: 2000-04-12
 PRIOR APPLICATION NUMBER: 60/256, 931
 PRIOR FILING DATE: 2000-12-21
 PRIOR APPLICATION NUMBER: 60/199, 384
 PRIOR FILING DATE: 2000-04-25
 NUMBER OF SEQ ID NOS: 2267
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 2022
 LENGTH: 264
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-833-245-2022

Query Match 1.1%; Score 8; DB 11; Length 264;

US-10-122-706-31
 ; Sequence 31, Application US/10122706
 ; Publication No. US20030119012A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Srinivasan, Maithreyan
 ; APPLICANT: Reifler, Michael
 ; TITLE OF INVENTION: Sulfurylase-luciferase Fusion Proteins
 ; FILE REFERENCE: 21465-504
 ; CURRENT APPLICATION NUMBER: US/10/122,706
 ; CURRENT FILING DATE: 2002-07-01
 ; PRIOR APPLICATION NUMBER: 60/335,949
 ; PRIOR FILING DATE: 2001-10-30
 ; NUMBER OF SEQ ID NOS: 31
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 31
 ; LENGTH: 309
 ; TYPE: PRT
 ; ORGANISM: Thermomonospora fusca
 US-10-122-706-31

Query Match 1.1%; Score 8; DB 14; Length 309;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 VLLFSGGK 286
 |||||
 Db 38 VLLFSGGK 45

RESULT 11
 US-09-833-245-770
 ; Sequence 770, Application US/09833245
 ; Publication No. US20040010134A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Human Genome Sciences, Inc.
 ; TITLE OF INVENTION: Albumin Fusion Proteins
 ; FILE REFERENCE: PF546PCT
 ; CURRENT APPLICATION NUMBER: US/09/833,245
 ; CURRENT FILING DATE: 2001-04-12
 ; PRIOR APPLICATION NUMBER: 60/229, 358
 ; PRIOR FILING DATE: 2000-04-12
 ; PRIOR APPLICATION NUMBER: 60/256, 931
 ; PRIOR FILING DATE: 2000-12-21
 ; PRIOR APPLICATION NUMBER: 60/199, 384
 ; PRIOR FILING DATE: 2000-04-25
 ; NUMBER OF SEQ ID NOS: 2267
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 770
 ; LENGTH: 383
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-833-245-770

Query Match 1.1%; Score 8; DB 11; Length 383;
 Best Local Similarity 100.0%; Pred. No. 92;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 PCQNGGQC 173
 |||||
 Db 100 PCQNGGQC 107

RESULT 12
 US-10-245-103-80
 ; Sequence 80, Application US/10245103
 ; Publication No. US20030068779A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin
 ; APPLICANT: Eaton, Dan
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Grimaldi, J. Christopher
 ; APPLICANT: Gurney, Austin

; APPLICANT: Smith, Victoria
 ; APPLICANT: Stephan, Jean-Phillippe
 ; APPLICANT: Watanabe, Colin
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin
 ; APPLICANT: Fong, Sherman
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; TITLE OF INVENTION: ACIDS ENCODING THE SAME
 ; FILE REFERENCE: P3630R1C112
 ; CURRENT APPLICATION NUMBER: US/10/245,103
 ; CURRENT FILING DATE: 2002-09-17
 ; PRIOR APPLICATION NUMBER: 10/197942
 ; PRIOR FILING DATE: 2002-07-18
 ; PRIOR APPLICATION NUMBER: 60/059114
 ; PRIOR FILING DATE: 1997-09-17
 ; PRIOR APPLICATION NUMBER: 60/063046
 ; PRIOR FILING DATE: 1997-10-24
 ; PRIOR APPLICATION NUMBER: 60/065027
 ; PRIOR FILING DATE: 1997-11-10
 ; PRIOR APPLICATION NUMBER: 60/079689
 ; PRIOR FILING DATE: 1998-03-27
 ; PRIOR APPLICATION NUMBER: 60/086478
 ; PRIOR FILING DATE: 1998-05-22
 ; PRIOR APPLICATION NUMBER: 60/087607
 ; PRIOR FILING DATE: 1998-06-02
 ; PRIOR APPLICATION NUMBER: 60/089801
 ; PRIOR FILING DATE: 1998-06-18
 ; PRIOR APPLICATION NUMBER: 60/090557
 ; PRIOR FILING DATE: 1998-06-24
 ; PRIOR APPLICATION NUMBER: 60/090689
 ; PRIOR FILING DATE: 1998-06-25
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 116
 ; SEQ ID NO 80
 ; LENGTH: 383
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-10-245-103-80

Query Match 1.1%; Score 8; DB 14; Length 383;
 Best Local Similarity 100.0%; Pred. No. 92;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 PCQNGGQC 173
 |||||
 Db 100 PCQNGGQC 107

RESULT 13
 US-10-245-107-80
 ; Sequence 80, Application US/10245107
 ; Publication No. US20030068779A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin
 ; APPLICANT: Eaton, Dan
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Grimaldi, J. Christopher
 ; APPLICANT: Gurney, Austin
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stephan, Jean-Phillippe
 ; APPLICANT: Watanabe, Colin
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin
 ; APPLICANT: Fong, Sherman
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; TITLE OF INVENTION: ACIDS ENCODING THE SAME
 ; FILE REFERENCE: P3630R1C71
 ; CURRENT APPLICATION NUMBER: US/10/245,107
 ; CURRENT FILING DATE: 2002-09-16
 ; PRIOR APPLICATION NUMBER: 10/197942
 ; PRIOR FILING DATE: 2002-07-18
 ; PRIOR APPLICATION NUMBER: 60/059114

;
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 80
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-245-107-80

Query Match 1.1%; Score 8; DB 14; Length 383;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 PCQNGGQC 173
Db 100 PCQNGGQC 107

RESULT 14
US-10-245-143-80
; Sequence 80, Application US/10245143
; Publication No. US20030068780A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630RIC90
; CURRENT APPLICATION NUMBER: US/10/245,143
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 80
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-245-143-80

Query Match 1.1%; Score 8; DB 14; Length 383;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

;
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 80
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-245-143-80

Query Match 1.1%; Score 8; DB 14; Length 383;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 PCQNGGQC 173
Db 100 PCQNGGQC 107

RESULT 15
US-10-245-771-80
; Sequence 80, Application US/10245771
; Publication No. US20030068781A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630RIC98
; CURRENT APPLICATION NUMBER: US/10/245,771
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 80
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-245-771-80

Query Match 1.1%; Score 8; DB 14; Length 383;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 166 PCONGQC 173
Db 100 PCONGQC 107

Search completed: March 9, 2004, 17:34:05
Job time : 52.7431 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: March 9, 2004, 17:23:39 ; Search time 24.1076 Seconds
(without alignments)
2813.016 Million cell updates/sec

Title: US-10-006-011A-3
Perfect score: 705
Sequence: 1 EIKITRPDSADGMILYNGQ.....QPLDLQRAQAGANTPCPS 705

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 78:.*
1: Piri:.*
2: Piri2:.*
3: Piri3:.*
4: Piri4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	705	100.0	4391	A38096	perlecan precursor
2	44	6.2	3707	S18252	heparan sulfate pr
3	9	1.3	861	A48825	Notch homolog Motc
4	9	1.3	1959	1 AGRT	agrin - rat
5	9	1.3	2531	2 S18188	notch protein homo
6	9	1.3	2531	2 A46019	notch-1 protein -
7	8	1.1	215	2 T48789	TFS1 related prote
8	8	1.1	332	2 A70772	probable sulfate a
9	8	1.1	430	2 A95335	conserved hypothet
10	8	1.1	497	2 AH2015	sodium/solute symp
11	8	1.1	500	2 F70012	leucyl aminopeptid
12	8	1.1	748	2 T00732	hypothetical prote
13	8	1.1	886	2 T39081	hypothetical prote
14	8	1.1	1203	2 A49175	Notch B protein -
15	8	1.1	1257	2 S28764	neurocan precursor
16	8	1.1	1258	2 S52781	neurocan - mouse
17	8	1.1	1328	2 T43060	agrin - electric r
18	8	1.1	1751	1 MMHUMH	laminin alpha-2 ch
19	8	1.1	1955	1 AGCH	agrin precursor -
20	8	1.1	2139	2 A35672	crumbs protein - f
21	8	1.1	2321	2 S78549	notch3 protein - h
22	8	1.1	2471	2 A49128	cell-fate determin
23	8	1.1	2531	2 T31070	notch homolog - se
24	8	1.1	3106	1 S53868	laminin alpha-2 ch
25	7	1.0	46	2 S21922	T-cell receptor al
26	7	1.0	93	2 S75571	hypothetical prote
27	7	1.0	96	2 E97765	mutator protein Mu
28	7	1.0	129	2 T29451	hypothetical prote
29	7	1.0	137	2 T43574	translocation prot

30	7	1.0	137	2 B40361	virC-region hypoth
31	7	1.0	139	2 S30967	gene 22 protein -
32	7	1.0	139	2 B95304	hypothetical prote
33	7	1.0	146	2 F75400	hypothetical prote
34	7	1.0	154	2 A99404	hypothetical prote
35	7	1.0	161	2 B84937	phosphotransferase
36	7	1.0	166	2 A82147	hypothetical prote
37	7	1.0	176	2 D71336	probable peptidyl-
38	7	1.0	179	2 A33104	chemotaxis protein
39	7	1.0	181	2 E87418	tenemrosin C - sea
40	7	1.0	196	2 G72718	probable transcrip
41	7	1.0	201	2 E75567	hypothetical prote
42	7	1.0	210	2 T03144	hypothetical prote
43	7	1.0	214	2 J24682	equinatoxin II pre
44	7	1.0	221	2 T24494	hypothetical prote
45	7	1.0	224	2 T10660	photosystem II pro

ALIGNMENTS

RESULT 1
A38096
perlecan precursor - human
N:Alternate names: Basement
C:Species: Homo sapiens (man)
C:Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #text change 05-Nov-1999
C:Accession: A38096; S19256; S77946; A41059; A40306; B33625; A33625; A41736
R:Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.
J. Biol. Chem. 267, 8544-8557, 1992
A:Title: Primary structure of the human heparan sulfate proteoglycan from basement membra
tor, laminin, neural cell adhesion molecules, and epidermal growth factor.
A:Reference number: A38096; MUID:92235084; PMID:1569102
A:Accession: A38096
A:Molecule type: mRNA
A:Residues: 1-4391 <MUR>
A:Cross-references: GB:M85289; NID:G184426; PIDN:AAA52700.1; PID:G184427
R:Kallunki, P.; Tryggvason, K.
J. Cell Biol. 116, 559-571, 1992
A:Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD prot
eoglycan core protein.
A:Reference number: A41736; MUID:92112994; PMID:1730768
A:Accession: S19256
A:Molecule type: mRNA
A:Residues: 1-57, 'D', '59-434, 'A', '436, 'FL', '438-449, 'Q', '451-502, 'A', '503-792, 'K', '794-908, 'R',
71-2979, 'H', '2981-2994, 'G', '2996-3167, 'T', '3169-3240, 'R', '3242-3426, 'R', '3428-3631, 'Q', '3633-3:
A:Cross-references: EMBL:X62515
R:Tryggvason, K.
submitted to the EMBL Data Library, October 1991
A:Reference number: S77946
A:Accession: S77946
A:Molecule type: mRNA
A:Residues: 1-57, 'D', '59-434, 'A', '436, 'FL', '438-449, 'Q', '451-502, 'A', '503-792, 'K', '794-908, 'R',
71-2979, 'H', '2981-2994, 'G', '2996-3167, 'T', '3169-3240, 'R', '3242-3426, 'R', '3428-3631, 'Q', '3633-4:
A:Cross-references: EMBL:X62515; NID:929469; PIDN:CAA44373.1; PID:929470
R:Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestilae, M.; Shows, T.B.; Tryggvason, K.
Genomics 11, 389-396, 1991
A:Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the g
A:Reference number: A41059; MUID:92120660; PMID:1685141
A:Accession: A41059
A:Molecule type: mRNA
A:Residues: 'R', '892-908, 'R', '910-1101, 'L', '1103-1132, 'L', '1134-1221, 'L', '1223-1397 <KA2>
A:Cross-references: GB:S76436; NID:9243370; PIDN:AAB21121.1; PID:9243371
R:Dodge, G.R.; Kovalevsky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo, F
Genomics 10, 673-680, 1991
A:Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellular
A:Reference number: A40306; MUID:91365376; PMID:1679749
A:Accession: A40306
A:Molecule type: mRNA
A:Residues: 1018-1405, 'G', '1407-1409, 'G', '1411-1465 <DOD>
A:Cross-references: GB:M64283; NID:G184424; PIDN:AAA52699.1; PID:G184425
R:Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Cassiman, J.J.; van den
J. Cell Biol. 109, 3199-3211, 1989

A>Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclonal
 A:Reference number: A33625; MUID:90078352; PMID:2687294
 A:Accession: B33625
 A:Molecule type: protein
 A:Residues: 1379-1384,'X',1386-1388,'X',1390-1398 <HE2>
 A:Accession: A33625
 A:Molecule type: protein
 A:Residues: 2166-2171,'X',2173-2175,'X',2177-2185 <HE3>
 A>Note: peptide potentially matches four different regions of sequence shown
 C:Genetics:
 A:Gene: GDB:HS9G2
 A:Cross-references: GDB:126372; OMIM:142461
 A:Map position: 1p36.1-1p36.1
 C:Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
 C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmembra
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:122-4391/Product: perlecan #status predicted <MAT>
 F:22-193/Domain: I <DOM1>
 F:194-530/Domain: II <DOM2>
 F:199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F:285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F:368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
 F:531-1676/Domain: III <DOM3>
 F:1159-1206/Domain: laminin-type EGF-like homology <LEG>
 F:1583-1610/Domain: laminin-type EGF-like homology <EGF>
 F:1613-1668/Domain: laminin-type EGF-like homology <EGF>
 F:1677-3686/Domain: IV <DOM4>
 F:2007-2034/Domain: transmembrane #status predicted <TRM>
 F:3687-4391/Domain: V <DOM5>
 F:3845-3880/Domain: EGF homology <EGF1>
 F:3888-3921/Domain: EGF homology <EGF>
 F:3953-4106/Domain: laminin G repeat homology <LG2>
 F:4147-4175/Domain: EGF homology <EGF2>
 F:4149-4151/Region: motor neuron attachment (L-R-E) motif
 F:4299-4301/Region: motor neuron attachment (L-R-E) motif
 F:65.71.76/Binding site: heparan sulfate (Ser) (covalent) #status predicted
 F:89.554,1755,2121,3072,3105,3279,3780,3836,4068/Binding site: carbohydrate
 F:2995,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 705; DB 2; Length 4391;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EIKTFPDSADGMLLYNGKRVPGSPNTLANRQPDFISFLVGRPERFEDAGSGWATI 60
 Db 3687 EIKTFPDSADGMLLYNGKRVPGSPNTLANRQPDFISFLVGRPERFEDAGSGWATI 3746
 QY 61 RHPTPLALGHFHTVTLRLSLTQSLIVGDLAPVNGTSQGRFQGLDLNEELYLGGYPDYGA 120
 Db 3747 RHPTPLALGHFHTVTLRLSLTQSLIVGDLAPVNGTSQGRFQGLDLNEELYLGGYPDYGA 3806
 QY 121 IPKAGLSGFTGCVRELRIOGEELVFHDLNLTAGHISHCTCRDRPCONGQCHDSSES 180
 Db 3807 IPKAGLSGFTGCVRELRIOGEELVFHDLNLTAGHISHCTCRDRPCONGQCHDSSES 3866
 QY 181 YVCVCPAGFTGSRCEHSQALCHPEACGPDATCNVRPDGRGYTCRCHLRSGLRCSEGYT 240
 Db 3867 YVCVCPAGFTGSRCEHSQALCHPEACGPDATCNVRPDGRGYTCRCHLRSGLRCSEGYT 3926
 QY 241 VTTSLGAGSYLALPALTTHHLRLDVEFKPLADGVLLFGSGSGPVEDFVSLAMVG 300
 Db 3927 VTTSLGAGSYLALPALTTHHLRLDVEFKPLADGVLLFGSGSGPVEDFVSLAMVG 3986
 QY 301 GHLEFRYELGSLAVLSASPLALGRWHRVSAERLNKDSLNRVNGRPLVLRSPGKSQGL 360
 Db 3987 GHLEFRYELGSLAVLSASPLALGRWHRVSAERLNKDSLNRVNGRPLVLRSPGKSQGL 4046
 QY 361 NLHTLLYLGVEPVPVLSPTATNSAHPRCGVGVSVNGKELDITYFLSGQIGQCYDSS 420
 Db 4047 NLHTLLYLGVEPVPVLSPTATNSAHPRCGVGVSVNGKELDITYFLSGQIGQCYDSS 4106

QY 421 PCERQPCQHGATCMPAGEYFQCLCRDGFKGLCEHENPCQRLRPFCLHGGTCQGRCLC 480
 Db 4107 PCERQPCQHGATCMPAGEYFQCLCRDGFKGLCEHENPCQRLRPFCLHGGTCQGRCLC 4166
 QY 481 LFGFSGPRCQCGSGHGAESDWHLESGSGNDAPGQYGFYFDDGFLAFPGHVFSSLSPEV 540
 Db 4167 LFGFSGPRCQCGSGHGAESDWHLESGSGNDAPGQYGFYFDDGFLAFPGHVFSSLSPEV 4226
 QY 541 PETIELEVTSTASGLLLWQGVGEAGGKDFISGLQDGHVFRYQLSGSEARLVSED 600
 Db 4227 PETIELEVTSTASGLLLWQGVGEAGGKDFISGLQDGHVFRYQLSGSEARLVSED 4286
 QY 601 PINDGEHRTVTLRREGGRGSIQVDEEELVSGRSPGNVAVNAKGSWYTCGAPDVATLTGG 660
 Db 4287 PINDGEHRTVTLRREGGRGSIQVDEEELVSGRSPGNVAVNAKGSWYTCGAPDVATLTGG 4346
 QY 661 RFSSGITGCVKLVLSHARPAPPPQPLDLQHRAGAGANTRPCPS 705
 Db 4347 RFSSGITGCVKLVLSHARPAPPPQPLDLQHRAGAGANTRPCPS 4391

RESULT 2
 S18252
 heparan sulfate proteoglycan - mouse
 N:Alternate names: perlecan
 C:Species: Mus musculus (house mouse)
 C>Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text_change 05-Nov-1999
 C:Accession: S18252; A31917; S66460
 R:Noonan, D.M.; Fuller, A.; Valente, P.; Cai, S.; Horigan, E.; Sasaki, M.; Yamada, Y.; Ha
 J. Biol. Chem. 266, 22939-22947, 1991
 A>Title: The complete sequence of perlecan, a basement membrane heparan sulfate proteogl
 adhesion molecule.
 A:Reference number: S18252; MUID:92078153; PMID:1744087
 A:Accession: S18252
 A:Molecule type: mRNA
 A:Residues: 1-3707 <NOO>
 A:Cross-references: EMBL:M77174; NID:g200295; PIDN:AAA39911.1; PID:g200296
 R:Noonan, D.M.; Horigan, E.A.; Ledbetter, S.R.; Vogeli, G.; Sasaki, M.; Yamada, Y.; Hass
 J. Biol. Chem. 263, 16379-16387, 1988
 U:Title: Identification of cDNA clones encoding different domains of the basement membra
 A:Reference number: A31917
 A:Accession: A31917
 A:Molecule type: mRNA
 A:Residues: 940-1601 <NO2>
 A:Cross-references: GB:J04054; NID:g200252; PIDN:AAA39899.1; PID:g200253
 A:Accession: B31917
 A:Molecule type: mRNA
 A:Residues: 1870-2600 <NO3>
 A:Cross-references: GB:J04055; NID:g200300; PIDN:AAA39912.1; PID:g200301
 R:Schulze, B.; Mann, K.; Battistutta, R.; Wiedemann, H.; Timpl, R.
 Eur. J. Biochem. 231, 551-556, 1995
 A:Title: Structural properties of recombinant domain III-3 of perlecan containing a glob
 A:Reference number: S66460; MUID:95377282; PMID:7649154
 A:Accession: S66460
 A:Molecule type: protein
 A:Residues: 1272-1274,'X',1276,'X',1278-1279 <SCH>
 C:Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
 C:Keywords: glycoprotein
 F:199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F:285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F:368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
 F:764-811/Domain: laminin-type EGF-like homology <LEG>
 F:1159-1206/Domain: laminin-type EGF-like homology <LEG7>
 F:1583-1610/Domain: laminin-type EGF-like homology <EGF>
 F:1613-1668/Domain: laminin-type EGF-like homology <EGF>
 F:3163-3198/Domain: laminin-type EGF-like homology <EGF>
 F:3270-3423/Domain: laminin G repeat homology <LG2>
 F:3464-3492/Domain: EGF homology <EGF7>
 F:1256,1891,2336,2394,2427/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 6.2%; Score 44; DB 2; Length 3707;
 Best Local Similarity 100.0%; Pred. No. 2,2e-36;

Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 SPTNLNRPQDFISGLVGGPPEFRFDAGSMATIRHTPTLALG 69
 DB 3030 SPTNLNRPQDFISGLVGGPPEFRFDAGSMATIRHTPTLALG 3073

RESULT 3
 A48825
 Notch homolog Notch protein - mouse (fragment)
 C:Species: Mus musculus (house mouse)
 C>Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Sep-2002
 C:Accession: A48825
 R:Reaume, A.G.; Conlon, R.A.; Zirngibl, R.; Yamaguchi, T.P.; Rossant, J.
 Dev. Biol. 154, 377-387, 1992
 A:Title: Expression analysis of a Notch homologue in the mouse embryo.
 A:Reference: A48825; MUID:93050801; PMID:1426644
 A:Accession: A48825
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-861 <REA>
 A:Experimental source: embryo
 A>Note: sequence extracted from NCBI backbone (NCBIP:119144)
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F:26-57/Domain: EGF homology <EGF>
 F:64-95/Domain: EGF homology <EGF>
 F:198-229/Domain: EGF homology <EGF2>
 F:441-472/Domain: EGF homology <EGX2>

Query Match 1.3%; Score 9; DB 2; Length 861;
 Best Local Similarity 100.0%; Pred. No. 2.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGGTCQ 474
 DB 240 PCLHGGTCQ 248

RESULT 4
 AGRT
 agrin - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
 C:Accession: JH0399; A38856
 R:Rupp, F.; Pavan, D.G.; Magill-Solc, C.; Cowan, D.M.; Scheller, R.H.
 Neuron 6, 811-823, 1991
 A:Title: Structure and expression of a rat agrin.
 A:Reference number: JH0399; MUID:91222570; PMID:1851019
 A:Accession: JH0399
 A:Molecule type: mRNA
 A:Residues: 1-1779; 1799-1959 <RUP>
 A:Cross-references: GB:M64780; NID:9202798; PIDN:AAA40703.1; PID:9202800
 A:Experimental source: embryonic spinal cord
 A>Note: it is uncertain whether Met-1, Met-18, or Met-24 is the initiator
 R:Rupp, F.; Oezcelik, T.; Linial, M.; Peterson, K.; Francke, U.; Scheller, R.
 J. Neurosci. 12, 3535-3544, 1992
 A:Title: Structure and chromosomal localization of the mammalian agrin gene.
 A:Reference number: A38856; MUID:92407628; PMID:1326508
 A:Accession: A38856
 A:Molecule type: mRNA
 A:Residues: 1780-1798 <RU2>
 A:Cross-references: GB:S44194
 C:Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine
 C:Comment: 90% of rat embryonic transcripts encode the variant labeled below as form 3.
 Ylcholine receptor clustering activity.
 C:Superfamily: agrin; EGF homology; Notch proteinase inhibitor homology; laminin G repeat
 C:Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
 F:1-1959/Product: agrin, form 1 #status predicted <AG1>
 F:1-1787,1799-1959/Product: agrin, form 4 #status predicted <AG4>
 F:1-1779,1799-1959/Product: agrin, form 3 #status predicted <AG3>
 F:1-1779,1788-1959/Product: agrin, form 5 #status predicted <AG5>
 F:1-1143,1153-1959/Product: agrin, form 2 #status predicted <AG2>
 F:22-50/Region: hydrophobic

F:98-137/Domain: Kazal proteinase inhibitor homology <KPI1>
 F:163-212/Domain: Kazal proteinase inhibitor homology <KPI2>
 F:236-284/Domain: Kazal proteinase inhibitor homology <KPI3>
 F:307-356/Domain: Kazal proteinase inhibitor homology <KPI4>
 F:381-429/Domain: Kazal proteinase inhibitor homology <KPI5>
 F:446-494/Domain: Kazal proteinase inhibitor homology <KPI6>
 F:511-559/Domain: Kazal proteinase inhibitor homology <KPI7>
 F:540-542/Region: motor neuron attachment (L-R-E) motif
 F:596-645/Domain: Kazal proteinase inhibitor homology <KPI8>
 F:688-739/Domain: laminin-type EGF-like homology <LE1>
 F:742-786/Domain: laminin-type EGF-like homology <LE2>
 F:814-864/Domain: Kazal proteinase inhibitor homology <KPI9>
 F:869-992/Region: serine/threonine-rich
 F:1084-1086/Region: motor neuron attachment (L-R-E) motif
 F:1147-1215/Region: EGF homology <EG1>
 F:1224-1257/Domain: EGF homology <EG2>
 F:1444-1476/Domain: EGF homology <EG3>
 F:1483-1515/Domain: EGF homology <EG4>
 F:1555-1706/Domain: EGF homology <EG5>
 F:1713-1747/Domain: EGF homology <EG6>
 F:1807-1959/Domain: laminin G repeat homology <LG3>
 F:97-116,105-137,171-191,180-212,244-263,252-284,316-335,324-356,389-408,397-429,454-473,
 -176,1483-1494,1488-1504,1506-1515/Disulfide bonds: #status predicted
 F:145,672,827,957/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.3%; Score 9; DB 1; Length 1959;
 Best Local Similarity 100.0%; Pred. No. 5.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGGTCQ 474
 DB 1228 PCLHGGTCQ 1236

RESULT 5
 S18188
 notch protein homolog - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 02-Aug-2002
 C:Accession: S18188
 R:Weinmaster, G.; Roberts, V.J.; Lemke, G.
 Development 113, 199-205, 1991
 A:Title: A homolog of Drosophila Notch expressed during mammalian development.
 A:Reference number: S18188; MUID:92111383; PMID:1764995
 A:Accession: S18188
 A:Molecule type: mRNA
 A:Residues: 1-2531 <WEI>
 A:Cross-references: EMBL:X57405; NID:957634; PID:957635
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F:987-1018/Domain: EGF homology <EGF1>
 F:1025-1056/Domain: EGF homology <EGF2>
 F:1233-1264/Domain: EGF homology <EGF3>
 F:1917-1949/Domain: ankyrin repeat homology <AN1>
 F:1950-1982/Domain: ankyrin repeat homology <AN2>
 F:1984-2016/Domain: ankyrin repeat homology <AN3>
 F:2017-2049/Domain: ankyrin repeat homology <AN4>
 F:2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match 1.3%; Score 9; DB 2; Length 2531;
 Best Local Similarity 100.0%; Pred. No. 6.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGGTCQ 474
 DB 1029 PCLHGGTCQ 1037

RESULT 6
 A46019
 notch-1 protein - mouse
 N:Alternate names: notch protein
 C:Species: Mus musculus (house mouse)

C>Date: 22-Sep-1993 #sequence revision 18-Nov-1994 #text change 07-Mar-2003
C/Accession: A46019; S25144; C49175; B46438; A46438; PH1539; S32109
R:del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; Grid
Genomics 15, 253-264, 1993
A>Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog of
A/Reference number: A46019; MUID:93194170; PMID:8449489
A/Accession: A46019
A/Status: not compared with conceptual translation
A/Molecule type: nucleic acid
A/Residues: 1-2531
A/Cross-references: GB:Z11886; GB:847228; NID:9288502; PIDN:CAA77941.1; PID:9288503
A/Note: sequence extracted from NCBI backbone (NCBIP:127318)
R:Francisco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J.;
submitted to the EMBL Data Library, April 1992
A/Description: Expression pattern of Notch, a mouse homolog of Drosophila Notch, suggest
A/Reference number: S25144
A/Accession: S25144
A/Molecule type: mRNA
A/Residues: 1551-2108, 'Q', 2110-2114, 'ALP', 2118-2170 <FRA>
A/Cross-references: EMBL:Z11886
R:Lardelli, M.; Lendahl, U.
Exp. Cell Res. 204, 364-372, 1993
A>Title: Notch A and Notch B--two mouse Notch homologues coexpressed in a wide variety of
A/Reference number: A49175; MUID:93178563; PMID:8440332
A/Accession: C49175
A/Status: preliminary; nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1161-1547 <LAR>
A/Cross-references: EMBL:X68278; NID:9287987; PIDN:CAA48339.1; PID:9287988
A/Experimental source: embryo
A/Note: sequence extracted from NCBI backbone (NCBIP:126159)
R:Kopan, R.; Weintraub, H.
J. Cell Biol. 121, 631-641, 1993
A>Title: Mouse Notch: expression in hair follicles correlates with cell fate determination
A/Reference number: A46438; MUID:93252998; PMID:8486742
A/Accession: B46438
A/Status: preliminary
A/Molecule type: nucleic acid
A/Residues: 1865-1932, 'RR', 1935-1937, 'L', 1938-1967, 'I', 1969-2044, 'IE', 2047-2052, 'S', 2054
A/Experimental source: embryo
A/Note: sequence extracted from NCBI backbone (NCBIP:131246; NCBIP:131247)
C/Comment: This protein has many EGF repeats and lin-12[1172]/Notch repeats.
C/Genetics: This protein is one of the neurogenic proteins controlling the decision between
A/Gene: notch-1
A/Map position: 2
A/Note: proximal region of chromosome 2
C/Superfamily: notch protein; ankyrin repeat homology; EGF homology
F:106-138/Domain: EGF homology <EGF1>
F:144-175/Domain: EGF homology <EGF1>
F:222-254/Domain: EGF homology <EGF2>
F:261-292/Domain: EGF homology <EGF2>
F:339-370/Domain: EGF homology <EGF3>
F:416-449/Domain: EGF homology <EGF3>
F:456-487/Domain: EGF homology <EGF4>
F:494-525/Domain: EGF homology <EGF4>
F:532-563/Domain: EGF homology <EGF5>
F:607-638/Domain: EGF homology <EGF5>
F:682-713/Domain: EGF homology <EGF6>
F:757-788/Domain: EGF homology <EGF6>
F:795-826/Domain: EGF homology <EGF7>
F:879-904/Domain: EGF homology <EGF7>
F:911-942/Domain: EGF homology <EGF8>
F:949-980/Domain: EGF homology <EGF8>
F:987-1018/Domain: EGF homology <EGF9>
F:1025-1056/Domain: EGF homology <EGF9>
F:1063-1094/Domain: EGF homology <EGF10>
F:1149-1180/Domain: EGF homology <EGF10>
F:1187-1218/Domain: EGF homology <EGF11>
F:1233-1264/Domain: EGF homology <EGF11>
F:1352-1383/Domain: EGF homology <EGF12>
F:1391-1425/Domain: EGF homology <EGF12>
F:1917-1948/Domain: ankyrin repeat homology <ANI>

F:1949-1981/Domain: ankyrin repeat homology <AN2>
F:1983-2015/Domain: ankyrin repeat homology <AN3>
F:2016-2048/Domain: ankyrin repeat homology <AN4>
F:2049-2081/Domain: ankyrin repeat homology <AN5>
Query Match 1.3%; Score 9; DB 2; Length 2531;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 466 PCLHGGTCC 474
DB 1029 PCLHGGTCC 1037
RESULT 7
T48789
TFS1 related protein [imported] - Neurospora crassa
N/Alternate names: protein 13E11.370
C/Species: Neurospora crassa
C/Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
C/Accession: T48789
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, April 2000
A/Reference number: Z24541
A/Accession: T48789
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-215 <SCH>
A/Cross-references: EMBL:AL333820; GSPDB:GN00112; NCSP:13E11.370
A/Experimental source: comid contig 13E11, strain 74
C/Genetics:
A/Gene: NCSP:13E11.370
A/Map position: 2
Query Match 1.1%; Score 8; DB 2; Length 215;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 90 LAPVNGTS 97
DB 136 LAPVNGTS 143
RESULT 8
A70772
Probable sulfate adenylate transferase subunit 2 - Mycobacterium tuberculosis (strain H3
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C/Accession: A70772
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: A70772
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-332 <COL>
A/Cross-references: GB:Z73419; GB:AL123456; NID:93261573; PIDN:CAA97751.1; PID:91322409
A/Experimental source: strain H37Rv
C/Genetics:
A/Gene: cysD
C/Superfamily: nodulation protein nodP
Query Match 1.1%; Score 8; DB 2; Length 332;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 279 VLLFSGGK 286
DB 64 VLLFSGGK 71

RESULT 9
A95935
conserved hypothetical membrane protein SMB21241 [imported] - *Sinorhizobium meliloti* (sc
C:Species: *Sinorhizobium meliloti*
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: A95935
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernar
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: A95935
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-430 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC49145.1; PID:gl5140630; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymB
R:Galibert, P.; Finan, T.M.; Long, S.R.; Fuhrer, A.; Abola, P.; Ampe, P.; Barloy-Hubler,
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SMB21241
A:Genome: plasmid

Query Match 1.1%; Score 8; DB 2; Length 430;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 SYLALPAL 258
|||||
DB 208 SYLALPAL 215

RESULT 10
AH2015
sodium/solute symporter [imported] - *Nostoc* sp. (strain PCC 7120)
C:Species: *Nostoc* sp. PCC 7120
A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AH2015
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shampo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing *Cyanobacterium Ana*
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2015
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-497 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA878044.1; PID:gl7135498; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all1678

Query Match 1.1%; Score 8; DB 2; Length 497;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 551 STASGLLL 558
|||||
DB 325 STASGLLL 332

RESULT 11
F70012
leucyl aminopeptidase homolog yuiE - *Bacillus subtilis*

C:Species: *Bacillus subtilis*
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 03-Mar-2003
C:Accession: F70012
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter,
C.; Bron, S.; Brouillet, S.; Bruchchi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerman, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferraz, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gallier,
iech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sator
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tseato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: F70012
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-500 <KUN>
A:Cross-references: GB:Z99120; GB:AL009126; NID:92635613; PIDN:CAB15195.1; PID:g2635702
A:Experimental source: strain 168
C:Genetics:
A:Gene: yuiE
C:Superfamily: Cytosol aminopeptidase

Query Match 1.1%; Score 8; DB 2; Length 500;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 653 DVATLTGG 660
|||||
DB 367 DVATLTGG 374

RESULT 12
T00732
hypothetical protein F22013.27 - *Arabidopsis thaliana*
C:Species: *Arabidopsis thaliana* (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 22-Oct-1999
C:Accession: T00732
R:Shinn, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.; Com
eologus, A.; Ecker, J.R.
submitted to the EMBL Data Library, April 1998
A:Description: Genomic sequence for *Arabidopsis thaliana* BAC F22013.
A:Reference number: Z14200
A:Accession: T00732
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-748 <SHI>
A:Cross-references: EMBL:AC003981; NID:93063438; PID:g3063465; GSPDB:GN000059; ATSP:F2201.
C:Genetics:
A:Gene: ATSP:F22013.27
A:Map position: 1
A:Introns: 329/3; 429/3; 460/2

Query Match 1.1%; Score 8; DB 2; Length 748;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 394 VSVNGKRL 401
|||||
DB 403 VSVNGKRL 410

RESULT 13
T39081
hypothetical protein SPAC7D4.03c - fission yeast (*Schizosaccharomyces pombe*)
C:Species: *Schizosaccharomyces pombe*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000

C/Accession: T39081
R/Gentiles, S.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1997.
A/Reference number: Z1826
A/Accession: T39081
A/Status: Preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-886 <GEN>
A/Cross-references: EMBL:Z99532; PIDN:CAB16720.1; GSPDB:GNO0066; SPDB:SPAC7D4.03c
A/Experimental source: strain 972h; cosmid c7D4
C/Genetics:
A/Gene: SPDB:SPAC7D4.03c
A/Map position: 1
C/Superfamily: Schizosaccharomyces pombe hypothetical protein SPAC7D4.03c

Query Match 1.1%; Score 8; DB 2; Length 886;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 394 VSVNGKRL 401
|||
DB 828 VSVNGKRL 835

RESULT 14
A49175
Motch B protein - mouse (fragment)
N/Alternate names: Notch homolog
C/Species: Mus musculus (house mouse)
C/Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 08-Sep-2002
C/Accession: A49175; PH1570; S32113
R/Lardelli, M.; Lendahl, U.
Exp. Cell Res. 204, 364-372, 1993
A/Title: Motch A and Motch B--two mouse Notch homologues coexpressed in a wide variety of tissues
A/Reference number: A49175; MUID:93178563; PMID:8440332
A/Accession: A49175
A/Status: preliminary; nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-1203 <LAR>
A/Cross-references: EMBL:X68279; NID:g287989; PIDN:CAA48340.1; PID:g287990
A/Experimental source: embryo
A/Note: sequence extracted from NCBI backbone (NCBIP:126158)
C/Comment: This protein has many EGF repeats and lin-12/Notch repeats.
C/Comment: This protein is one of the neurogenic proteins controlling the decision between
C/Superfamily: notch protein; ankyrin repeat homology; EGF homology
F:143-174/Domain: EGF homology <EGX1>
F:482-513/Domain: EGF homology <EGF1>
F:560-591/Domain: EGF homology <EGF>
F:674-705/Domain: EGF homology <EGX2>
F:712-743/Domain: EGF homology <EGF3>
F:836-867/Domain: EGF homology <EGX3>

Query Match 1.1%; Score 8; DB 2; Length 1203;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 426 PCQHGATC 433
|||
DB 840 PCQHGATC 847

RESULT 15
S28764
neurocan precursor - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 22-Nov-1993 #sequence_revision 01-Sep-1995 #text_change 04-Feb-2000
C/Accession: S28764
R/Rauch, U.; Karthikeyan, L.; Maurel, P.; Margolis, R.U.; Margolis, R.K.
J. Biol. Chem. 267, 19536-19547, 1992
A/Title: Cloning and primary structure of neurocan, a developmentally regulated, aggrega
A/Reference number: S28764; MUID:92406907; PMID:1326557
A/Accession: S28764
A/Molecule type: mRNA

A/Residues: 1-1257 <RAU>
A/Cross-references: EMBL:M97161; NID:g205649; PIDN:AAC37679.1; PID:g205650
C/Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EG
C/Keywords: chondroitin sulfate proteoglycan; glycoprotein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-1257/Product: neurocan #status predicted <MAT>
F:176-253/Domain: link protein repeat homology <LNK1>
F:274-355/Domain: link protein repeat homology <LNK1>
F:364-366/Region: cell attachment (R-G-D) motif
F:953-984/Domain: EGF homology <EGF>
F:1029-1149/Domain: C-type lectin homology <LCH>
F:1156-1212/Domain: complement factor H repeat homology <FHD>
F:121-339,737,967,1164/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:372-410/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
F:944/Binding site: chondroitin sulfate (Ser) (covalent) #status experimental

Query Match 1.1%; Score 8; DB 2; Length 1257;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGATC 473
|||
DB 957 PCLHGATC 964

Search completed: March 9, 2004, 17:31:26
Job time: 26.1076 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:21:48 ; Search time 17.0517 Seconds
(without alignments)

2152.832 Million cell updates/sec

Title: US-10-006-011a-3

Perfect score: 705

Sequence: 1 EIKITFRDSADGMLYNGQ.....QPLDLQHRAGANTRPCPS 705

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	705	100.0	4391	1	PGBM_HUMAN
2	44	6.2	3707	1	PGBM_MOUSE
3	9	1.3	1959	1	AGRI_RAT
4	9	1.3	2531	1	NTC1_MOUSE
5	9	1.3	2531	1	NTC1_RAT
6	8	1.1	309	1	CYSD_MYCTU
7	8	1.1	500	1	AMPA_BACSU
8	8	1.1	1257	1	PGCN_RAT
9	8	1.1	1268	1	PGCN_MOUSE
10	8	1.1	1321	1	PGCN_HUMAN
11	8	1.1	1328	1	AGRI_DISOM
12	8	1.1	1955	1	AGRI_CHICK
13	8	1.1	2139	1	CRB_DROME
14	8	1.1	2321	1	NTC3_HUMAN
15	8	1.1	2470	1	NTC2_MOUSE
16	8	1.1	2471	1	NTC2_HUMAN
17	8	1.1	2471	1	NTC2_RAT
18	8	1.1	3106	1	LM42_MOUSE
19	8	1.1	3110	1	LM42_HUMAN
20	7	1.0	78	1	Y009_BFL2
21	7	1.0	137	1	YSCB_YEREN
22	7	1.0	139	1	VG22_BM15
23	7	1.0	161	1	PTGA_BUCAL
24	7	1.0	176	1	SLYD_TREPA
25	7	1.0	184	1	RS7_TREAC
26	7	1.0	184	1	RS7_THEVO
27	7	1.0	192	1	RR4B_CVACA
28	7	1.0	214	1	EOT2_ACTEQ
29	7	1.0	214	1	EOT4_ACTEQ
30	7	1.0	214	1	EOT5_ACTEQ
31	7	1.0	229	1	MYB1_NEUCR
32	7	1.0	240	1	YINC_STRLI
33	7	1.0	244	1	YR12_CABEL

34 7 1.0 248 1 SCTM_HUMAN
35 7 1.0 250 1 VL01_VACCC
36 7 1.0 250 1 VL01_VACCV
37 7 1.0 250 1 VL01_VARV
38 7 1.0 255 1 TPDC_ALCEU
39 7 1.0 273 1 RL2_TREPA
40 7 1.0 277 1 CHIS_BACSU
41 7 1.0 296 1 C1CG_RHOSH
42 7 1.0 297 1 DHSB_DROME
43 7 1.0 305 1 FMT_THETH
44 7 1.0 320 1 MDCF_RHIME
45 7 1.0 325 1 NAS2_ORISA

ALIGNMENTS

RESULT 1
PGBM_HUMAN
ID PGBM_HUMAN STANDARD; PRT; 4391 AA.
AC P98160; Q16287; Q9H3V5;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (perlecan) (PLC).
GN HSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112994; PubMed=1730769;
RA Kallunki P., Tryggvason K.;
RT "Human basement membrane heparan sulfate proteoglycan core protein: a
RT 467-kD protein containing multiple domains resembling elements of the
RT low density lipoprotein receptor, laminin, neural cell adhesion
RT molecules, and epidermal growth factor.";
RL J. Cell Biol. 116:559-571(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Colon, and Skin;
TX MEDLINE=92235084; PubMed=1569102;
RA Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT "Primary structure of the human heparan sulfate proteoglycan from
RT basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT domains homologous to the low density lipoprotein receptor, laminin,
RT neural cell adhesion molecules, and epidermal growth factor.";
RL J. Biol. Chem. 267:8544-8557(1992).
RN [3]
RP SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SJS1 TYR-1532.
RX MEDLINE=20553141; PubMed=1101850;
RA Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barral D.,
RA Beighton P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,
RA Samson D., Utziberea J.A., Lehmann-Horn F., Weissenbach J.,
RA Hentati F., Fontaine B.;
RT "Perlecan, the major proteoglycan of basement membranes, is altered in
RT patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia).";
RL Nat. Genet. 26:480-483(2000).
RN [4]
RP SEQUENCE OF 1016-1470 FROM N.A.
RX TISSUE=Colon;
TX MEDLINE=91365376; PubMed=1679749;
RA Dodge G.R., Kovalsky I., Chu M.L., Hassell J.R., McBride O.W.,
RA Yi H.F., Iozzo R.V.;
RT "Heparan sulfate proteoglycan of human colon: partial molecular
RT cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT short arm of human chromosome 1.";
RL Genomics 10:673-680(1991).
RN [5]
RP SEQUENCE OF 890-1396 FROM N.A.
RX TISSUE=Fibrosarcoma;

RX MEDLINE=92120660; PubMed=1685141;
RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA Tryggvason K.;
RT "Cloning of human heparan sulfate proteoglycan core protein,
RT assignment of the gene (HSPG2) to lp36.1--p35 and identification of
RT a BamHI restriction fragment length polymorphism.";
RL Genomics 11:389-396(1991).
RN [6]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=94052171; PubMed=8234307;
RA Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RA "Structural characterization of the complete human perlecan gene and
RT its promoter.";
RT Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
RN [7]
RP CARBOHYDRATE-LINKAGE SITE ASN-2121.
RX MEDLINE=22660472; PubMed=12754519;
RA Zhang H., Li X.-J., Martin D.B., Aebbersold R.;
RA "Identification and quantification of N-linked glycoproteins using
RT hydrazide chemistry, stable isotope labeling and mass spectrometry.";
RL Nat. Biotechnol. 21:660-666(2003).
CC -!- FUNCTION: This protein is an integral component of basement
CC membranes. It is responsible for the fixed negative electrostatic
CC charge and is involved in the charge-selective ultrafiltration
CC properties. It serves as an attachment substrate for cells.
CC -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC dimers or stellate structures. It interacts with other basement
CC membrane components such as laminin, prolargin and collagen type
CC IV.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Found in the basement membranes.
CC -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -!- DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel
CC syndrome (SJS1) (MIM:255800); a rare autosomal recessive disorder
CC characterized by permanent myotonia (prolonged failure of muscle
CC relaxation) and skeletal dysplasia, resulting in reduced stature,
CC kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.
CC -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC -!- SIMILARITY: Contains 11 laminin EGF-like domains.
CC -!- SIMILARITY: Contains 3 laminin IV domains.
CC -!- SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 3 laminin G-like domains.
CC -!- SIMILARITY: Contains 4 EGF-like domains.
CC -!- SIMILARITY: Contains 1 SEA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X62515; CAA44373.1; -;
CC EMBL; M85289; ABA52700.1; -;
CC EMBL; AL445795; CAC18534.1; -;
CC EMBL; M64283; ABA52699.1; -;
CC EMBL; S76436; AAB21121.2; -;
CC EMBL; L22078; -; NOT ANNOTATED_CDS.
CC PIR; A38096; A38096_-;
CC HSPG; P00740; 1EDM.
CC Siena-2DPAGE; P98160; -;
CC Genew; HGNC:5273; HSPG2.
CC MIM; 142461; -;
CC MIM; 255800; -;
CC InterPro; IPR008985; ConA_like_lec_g1.
CC InterPro; IPR000742; EGF_2.
CC InterPro; IPR006209; EGF-like.
CC InterPro; IPR006210; IEGF.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003599; Ig.
CC InterPro; IPR003598; Ig_c2.

DR InterPro; IPR003596; Ig_v.
DR InterPro; IPR000034; Laminin B.
DR InterPro; IPR002049; Laminin EGF.
DR InterPro; IPR001791; Laminin G.
DR InterPro; IPR002172; LDL receptor_A.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF00008; EGF_4.
DR Pfam; PF00047; ig_22.
DR Pfam; PF00052; laminin_B_3.
DR Pfam; PF00053; laminin_EGF_7.
DR Pfam; PF00054; laminin_G_3.
DR Pfam; PF00057; ldl_recept_a_4.
DR Pfam; PF01390; SEA_1.
DR PRINTS; PR00261; LDLRECEPTOR.
DR ProDom; PD003031; Laminin_B_3.
DR SMART; SM00181; EGF_15.
DR SMART; SM00180; EGF_Lam; 12.
DR SMART; SM00409; IG_22.
DR SMART; SM00408; IGC2; 21.
DR SMART; SM00406; IGV_7.
DR SMART; SM00281; LamB_3.
DR SMART; SM00282; LamG_3.
DR SMART; SM00192; LDLa_4.
DR SMART; SM0200; SEA_1.
DR PROSITE; PS00022; EGF_1; 9.
DR PROSITE; PS01186; EGF_2; 6.
DR PROSITE; PS00026; EGF_3; 4.
DR PROSITE; PS00835; IG_LIKE; 22.
DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR PROSITE; PS01209; LDLRA_1; 4.
DR PROSITE; PS00688; LDLRA_2; 4.
DR PROSITE; PS00204; SEA; 1.
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
KW Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW Extracellular matrix; EGF-like domain; Disease mutation.
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2437 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 324 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 501 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865

Query Match 100.0%; Score 705; DB 1; Length 4391;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EIKITFPDSADGMLLYNGKRVGSPSTNLNARQDPFISFGLVGRPFERFDAGSGMATI 60
DB 3687 EIKITFPDSADGMLLYNGKRVGSPSTNLNARQDPFISFGLVGRPFERFDAGSGMATI 3746

QY 61 RHPTPLALGHFTVTLRSITQSLVGLDAPVNGTSQKFGQLDNEELYIGGYDYGA 120
DB 3747 RHPTPLALGHFTVTLRSITQSLVGLDAPVNGTSQKFGQLDNEELYIGGYDYGA 3806

QY 121 IPKAGLSSGFIGVRELIQGEIVFHDNLNLTAGHISHCPTCRDRPCQNGQCHDESSS 180
DB 3807 IPKAGLSSGFIGVRELIQGEIVFHDNLNLTAGHISHCPTCRDRPCQNGQCHDESSS 3866

QY 181 YVCVCPAGFTGSCHEQSALHCHPEACGPDATCNRPDGRGYTCRCHLGRSLRCEEGVT 240
DB 3867 YVCVCPAGFTGSCHEQSALHCHPEACGPDATCNRPDGRGYTCRCHLGRSLRCEEGVT 3926

QY 241 VTPFSLSGAGSYLALPALNTNTHLRLDVEFKPLADGVLLFSGKSGPVEDFVSLAMWG 300
DB 3927 VTPFSLSGAGSYLALPALNTNTHLRLDVEFKPLADGVLLFSGKSGPVEDFVSLAMWG 3986

QY 301 GHLEFRLYELGSLAVLRSAPLALGRHVRVSAERLNKOGSLRNGRPVLRSSPKSQSL 360
DB 3987 GHLEFRLYELGSLAVLRSAPLALGRHVRVSAERLNKOGSLRNGRPVLRSSPKSQSL 4046

QY 361 NLHTLLYLGVEVPSVPLSPATNMSAHRFCVGVSVNGKRLDLYFSLGSGIGQCYDSS 420
DB 4047 NLHTLLYLGVEVPSVPLSPATNMSAHRFCVGVSVNGKRLDLYFSLGSGIGQCYDSS 4106

QY 421 PCERQPCQHGATCMWAGEYFQCLRDGFKGLDCEHENPCQLRPFCLHGTCQGTCLIC 480
DB 4107 PCERQPCQHGATCMWAGEYFQCLRDGFKGLDCEHENPCQLRPFCLHGTCQGTCLIC 4166

QY 481 LPFGSPRCQSGHGIASDWHLESGGNDAPGQYGFYHDDGLAPGHVFSLSLEV 540
DB 4167 LPFGSPRCQSGHGIASDWHLESGGNDAPGQYGFYHDDGLAPGHVFSLSLEV 4226

QY 541 PETIELEVTSTAGLLWQGVVEGAGQKDFISGLQDGHVFRYQLGSGEARLVSED 600
DB 4227 PETIELEVTSTAGLLWQGVVEGAGQKDFISGLQDGHVFRYQLGSGEARLVSED 4286

QY 601 PINDGEWHVTLREGRGSIQVDGELVSGRSPGNVAVNAKGSVYTGAPDVATLTGG 660
DB 4287 PINDGEWHVTLREGRGSIQVDGELVSGRSPGNVAVNAKGSVYTGAPDVATLTGG 4346

QY 661 RFSSGITGCVKNLVLSARPPPPQPLDQRAAGANTRPCPS 705
DB 4347 RFSSGITGCVKNLVLSARPPPPQPLDQRAAGANTRPCPS 4391

RESULT 2

PGEM_MOUSE STANDARD; PRT; 3707 AA.
ID PGEM_MOUSE
AC Q05753;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (Perlecan) (PLC).
GN HSPG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=92078153; PubMed=1744087;
RA Noonan D.M., Fullie A., Valente P., Cai S., Horigan E., Sasaki M.,

Yamada Y., Hassell J.R.;
"The complete sequence of perlecan, a basement membrane heparan sulfate proteoglycan, reveals extensive similarity with laminin A chain, low density lipoprotein-receptor, and the neural cell adhesion molecule.";
J. Biol. Chem. 266:22939-22947(1991).
[2]
SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89034110; PubMed=2972708;
RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogeli G., Sasaki M., Yamada Y., Hassell J.R.;
"Identification of cDNA clones encoding different domains of the basement membrane heparan sulfate proteoglycan.";
J. Biol. Chem. 263:16379-16387(1988).
CC -!- FUNCTION: This protein is an integral component of basement membranes. It is responsible for the fixed negative electrostatic charge and is involved in the charge-selective ultrafiltration properties. It serves as an attachment substrate for cells.
CC -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in dimers or stellate structures. It interacts with other basement membrane components such as laminin, prolargin and collagen type IV.
CC IV.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Found in the basement membranes.
CC -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC -!- SIMILARITY: Contains 11 laminin EGF-like domains.
CC -!- SIMILARITY: Contains 3 laminin IV domains.
CC -!- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 3 laminin G-like domains.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 1 SEA domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M7174; AAA39911.1; -;
EMBL; J04054; AAA39899.1; -;
EMBL; J04055; AAA39912.1; -;
PIR; S18252; S18252.
PDB; 1GL4; 28-NOV-01.
MGD; MGI:96257; HSPG2.
GO; GO:0005604; C:basement membrane; IDA.
GO; GO:0008104; P:protein localization; IMP.
InterPro; IPR006985; ConA_like_lec_gl.
InterPro; IPR000742; EGF_2.
InterPro; IPR006209; EGF-like.
InterPro; IPR007110; Ig-like.
InterPro; IPR003598; Ig_c2.
InterPro; IPR000034; Laminin_B.
InterPro; IPR002049; Laminin_EGF.
InterPro; IPR001791; Laminin_G.
InterPro; IPR002172; LDL_receptor_A.
InterPro; IPR000082; SEA_domain.
Pfam; PF000008; EGF; 4.
Pfam; PF000047; Ig; 15.
Pfam; PF000052; laminin_B; 3.
Pfam; PF000053; laminin_EGF; 7.
Pfam; PF000054; laminin_G; 3.
Pfam; PF000057; ldl_recept_a; 4.
Pfam; PF01390; SEA; 1.
PRINTS; PR00261; LDLRECEPTOR.
ProDom; PD003031; Laminin_B; 3.
SMART; SM00180; EGF_Lam; 7.
SMART; SM00408; IGC2; 14.
SMART; SM00281; LamB; 3.
SMART; SM00282; LamG; 3.

DR SMART; SMO0192; LDLA; 4.
 DR SMART; SMO0200; SEA; 1.
 DR PROSITE; PS00022; EGF_1; 8.
 DR PROSITE; PS01186; EGF_2; 5.
 DR PROSITE; PS00026; EGF_3; 4.
 DR PROSITE; PS00835; IG Like; 15.
 DR PROSITE; PS00025; LAM G DOMAIN; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
 DR PROSITE; PS01209; LDLRA_1; 4.
 DR PROSITE; PS00068; LDLRA_2; 4.
 DR PROSITE; PS00024; SEA; 1.
 KW Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
 KW Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
 KW Extracellular matrix; EGF-like domain; 3D-structure.
 FT SIGNAL 1 21
 FT CHAIN 22 3707
 FT DOMAIN 80 194
 FT DOMAIN 195 234
 FT DOMAIN 281 319
 FT DOMAIN 320 359
 FT DOMAIN 360 403
 FT DOMAIN 404 504
 FT DOMAIN 521 530
 FT DOMAIN 531 730
 FT DOMAIN 731 763
 FT DOMAIN 764 813
 FT DOMAIN 814 871
 FT DOMAIN 879 923
 FT DOMAIN 924 933
 FT DOMAIN 934 1125
 FT DOMAIN 1126 1158
 FT DOMAIN 1159 1208
 FT DOMAIN 1209 1265
 FT DOMAIN 1275 1324
 FT DOMAIN 1325 1334
 FT DOMAIN 1335 1529
 FT DOMAIN 1563 1612
 FT DOMAIN 1613 1670
 FT DOMAIN 1677 1771
 FT DOMAIN 1772 1865
 FT DOMAIN 1866 1954
 FT DOMAIN 1955 2049
 FT DOMAIN 2050 2149
 FT DOMAIN 2149 2244
 FT DOMAIN 2245 2343
 FT DOMAIN 2344 2436
 FT DOMAIN 2437 2532
 FT DOMAIN 2533 2619
 FT DOMAIN 2620 2720
 FT DOMAIN 2721 2809
 FT DOMAIN 2810 2895
 FT DOMAIN 2896 2980
 FT DOMAIN 2984 3162
 FT DOMAIN 3163 3241
 FT DOMAIN 3245 3425
 FT DOMAIN 3518 3705
 FT SITE 65 67
 FT SITE 71 73
 FT SITE 76 78
 FT SITE 3615 3617
 FT DISULFID 199 212
 FT DISULFID 206 225
 FT DISULFID 219 234
 FT DISULFID 285 297
 FT DISULFID 292 310
 FT DISULFID 304 319
 FT DISULFID 325 337
 FT DISULFID 332 350
 FT DISULFID 344 359
 FT DISULFID 368 381

FT DISULFID 375 394
 FT DISULFID 388 403
 FT DISULFID 428 479
 FT DISULFID 764 773
 FT DISULFID 766 780
 FT DISULFID 783 792
 FT DISULFID 795 811
 FT DISULFID 814 829
 FT DISULFID 816 839
 FT DISULFID 842 851
 FT DISULFID 854 869
 FT DISULFID 1159 1168
 FT DISULFID 1161 1175
 FT DISULFID 1178 1187
 FT DISULFID 1190 1206
 FT DISULFID 1209 1224
 FT DISULFID 1211 1234
 FT DISULFID 1237 1246
 FT DISULFID 1249 1263
 FT DISULFID 1275 1287
 FT DISULFID 1277 1293
 FT DISULFID 1295 1304
 FT DISULFID 1307 1322
 FT DISULFID 1563 1572
 FT DISULFID 1565 1579
 FT DISULFID 1582 1591
 FT DISULFID 1594 1610
 FT DISULFID 1613 1628
 FT DISULFID 1615 1638
 FT DISULFID 1641 1650
 FT DISULFID 1653 1668
 FT DISULFID 1792 1839
 FT DISULFID 1886 1932
 FT DISULFID 1976 2021
 FT DISULFID 2073 2118
 FT DISULFID 2170 2215
 FT DISULFID 2268 2313
 Query Match 6.2%; Score 44; DB 1; Length 3707;
 Best Local Similarity 100.0%; Pred.No. 4.6e-35;
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 26 SPTNLNRRQDFISFGLVGGPRFDFDAGSGMATIRHPTPLALG 69
 DB 3030 SPTNLNRRQDFISFGLVGGPRFDFDAGSGMATIRHPTPLALG 3073
 RESULT 3
 AGRI_RAT
 ID AGRI_RAT STANDARD; PRT; 1959 AA.
 AC P25304; Q63034;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Agrin precursor.
 GN AGRI.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE OF 1-1779 AND 1799-1959 FROM N.A.
 RC TISSUE=Embryonic spinal cord;
 RX MEDLINE=9122570; PubMed=1851019;
 RA Rupp F., Payan D.G., Megill-Solc C., Cowan D.M., Scheller R.H.;
 RT "Structure and expression of a rat agrin."
 RL Neuron 6:811-823(1991).
 RN [2]
 RP SEQUENCE OF 1777-1801 FROM N.A.
 RX MEDLINE=92407628; PubMed=1326608;
 RA Rupp F., Oezcelik T., Linial M., Peterson K., Francke U., Scheller R.;
 RT "Structure and chromosomal localization of the mammalian agrin gene."
 RL J. Neurosci. 12:3535-3544(1992).

FT	DOMAIN	65	137	KAZAL-LIKE 1.
FT	DOMAIN	141	212	KAZAL-LIKE 2.
FT	DOMAIN	213	284	KAZAL-LIKE 3.
FT	DOMAIN	287	356	KAZAL-LIKE 4.
FT	DOMAIN	361	429	KAZAL-LIKE 5.
FT	DOMAIN	430	494	KAZAL-LIKE 6.
FT	DOMAIN	495	559	KAZAL-LIKE 7.
FT	DOMAIN	563	645	KAZAL-LIKE 8.
FT	DOMAIN	688	741	LAMININ EGF-LIKE 1.
FT	DOMAIN	742	788	LAMININ EGF-LIKE 2.
FT	DOMAIN	794	864	KAZAL-LIKE 9.
FT	DOMAIN	1023	1145	EGF-LIKE 1.
FT	DOMAIN	1220	1258	LAMININ G-LIKE 1.
FT	DOMAIN	1263	1439	EGF-LIKE 2.
FT	DOMAIN	1440	1477	EGF-LIKE 3.
FT	DOMAIN	1479	1516	EGF-LIKE 4.
FT	DOMAIN	1526	1708	LAMININ G-LIKE 2.
FT	DOMAIN	1709	1748	EGF-LIKE 4.
FT	DOMAIN	1784	1956	LAMININ G-LIKE 3.
FT	DOMAIN	869	992	SER/THR-RICH.
FT	DOMAIN	1147	1215	SER/THR-RICH.
FT	DISULFID	97	116	POTENTIAL.
FT	DISULFID	105	137	POTENTIAL.
FT	DISULFID	171	191	POTENTIAL.
FT	DISULFID	180	212	POTENTIAL.
FT	DISULFID	244	263	POTENTIAL.
FT	DISULFID	252	284	POTENTIAL.
FT	DISULFID	316	335	POTENTIAL.
FT	DISULFID	324	356	POTENTIAL.
FT	DISULFID	389	408	POTENTIAL.
FT	DISULFID	397	429	POTENTIAL.
FT	DISULFID	454	473	POTENTIAL.
FT	DISULFID	462	494	POTENTIAL.
FT	DISULFID	518	538	POTENTIAL.
FT	DISULFID	527	559	POTENTIAL.
FT	DISULFID	604	624	POTENTIAL.
FT	DISULFID	613	645	POTENTIAL.
FT	DISULFID	688	700	BY SIMILARITY.
FT	DISULFID	690	707	BY SIMILARITY.
FT	DISULFID	709	718	BY SIMILARITY.
FT	DISULFID	721	739	BY SIMILARITY.
FT	DISULFID	742	754	BY SIMILARITY.
FT	DISULFID	744	761	BY SIMILARITY.
FT	DISULFID	763	772	BY SIMILARITY.
FT	DISULFID	775	786	BY SIMILARITY.
FT	DISULFID	823	843	POTENTIAL.
FT	DISULFID	832	864	POTENTIAL.
FT	DISULFID	1224	1235	BY SIMILARITY.
FT	DISULFID	1229	1246	BY SIMILARITY.
FT	DISULFID	1248	1257	BY SIMILARITY.
FT	DISULFID	1444	1455	POTENTIAL.
FT	DISULFID	1449	1465	POTENTIAL.
FT	DISULFID	1467	1476	POTENTIAL.
FT	DISULFID	1483	1494	BY SIMILARITY.
FT	DISULFID	1488	1504	BY SIMILARITY.
FT	DISULFID	1506	1515	BY SIMILARITY.
FT	DISULFID	1713	1727	BY SIMILARITY.
FT	DISULFID	1721	1736	BY SIMILARITY.
FT	DISULFID	1738	1747	BY SIMILARITY.
FT	CARBOHYD	145	145	N-LINKED (GLCNAC. .)
FT	CARBOHYD	672	672	N-LINKED (GLCNAC. .)
FT	CARBOHYD	827	827	N-LINKED (GLCNAC. .)
FT	CARBOHYD	957	957	N-LINKED (GLCNAC. .)
FT	VARSPLIC	1144	1152	Missing (in isoform 2).
FT	FT			/FTid=VSP_001365.
FT	FT			Missing (in isoform 3).
FT	FT			/FTid=VSP_001366.
FT	FT			Missing (in isoform 4).
FT	FT			/FTid=VSP_001367.
FT	FT			Missing (in isoform 5).
FT	FT			/FTid=VSP_001368.
FT	FT			Missing (in isoform 2).
FT	FT			/FTid=VSP_001365.
FT	FT			Missing (in isoform 3).
FT	FT			/FTid=VSP_001366.
FT	FT			Missing (in isoform 4).
FT	FT			/FTid=VSP_001367.
FT	FT			Missing (in isoform 5).
FT	FT			/FTid=VSP_001368.
FT	FT			Missing (in isoform 2).
FT	FT			/FTid=VSP_001365.
FT	FT			Missing (in isoform 3).
FT	FT			/FTid=VSP_001366.
FT	FT			Missing (in isoform 4).
FT	FT			/FTid=VSP_001367.
FT	FT			Missing (in isoform 5).
FT	FT			/FTid=VSP_001368.
FT	FT			Missing (in isoform 2).
FT	FT			/FTid=VSP_001365.
FT	FT			Missing (in isoform 3).
FT	FT			/FTid=VSP_001366.
FT	FT			Missing (in isoform 4).
FT	FT			/FTid=VSP_001367.
FT	FT			Missing (in isoform 5).

Query Match 1.3%; Score 9, DB 1; Length 1959;
 Best Local Similarity 100.0%; Pred. No. 4;
 Matches 9; Conservative 0; Mismatches 0; Gaps 0;

QY 466 PCLHGGTCQ 474
 DB 1228 PCLHGGTCQ 1236

RESULT 4
 NTC1 MOUSE
 ID NTC1 MOUSE STANDARD; PRT: 2531 AA.
 AC Q01705; Q06007; Q61905; Q99JC2; Q9QW58; Q9ROX7;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (Notch A) (mrl4) (p300).
 GN NOTCH1 OR NOTCH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Embryo;
 RX MEDLINE=93194170; PubMed=8449489;
 RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
 RA Copeland N.G., Gridley T.;
 RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
 RT homolog of Drosophila Notch.";
 RL Homologs 15:259-264 (1993).
 CC [2]
 RP SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
 RC STRAIN=CD-1; TISSUE=Embryo;
 RX MEDLINE=93050801; PubMed=1426644;
 RA Reaume A.G., Conlon R.A., Ziringibl R., Yamaguchi T.P., Rossant J.;
 RA "Expression, analysis of a Notch homologue in the mouse embryo.";
 RT Dev. Biol. 154:377-387 (1992).
 CC [3]
 RP SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
 RC TISSUE=Embryo;
 RX MEDLINE=93048835; PubMed=1425352;
 RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
 RA Greenpan R.J., McMahon A.P., Gridley T.;
 RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,
 RT suggests an important role in early postimplantation mouse
 RT development.";
 RL Development 115:737-744 (1992).
 CC [4]
 RP SEQUENCE OF 1161-1547 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
 RX MEDLINE=93178563; PubMed=8440332;
 RA Lardelli M., Lendahl U.;
 RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
 RT wide variety of tissues.";
 RL Exp. Cell Res. 204:364-372 (1993).
 CC [5]
 RP SEQUENCE OF 1659-1673 FROM N.A.
 RX MEDLINE=93064499; PubMed=10437788;
 RA Lee J.S., Isimoto A., Yanagawa S.I.;
 RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads
 RT to induction of HES-1 in a mouse T lymphoma cell line, DL-3.";
 RL FEBS Lett. 455:276-280 (1999).
 CC [6]
 RP SEQUENCE OF 1950-2201 FROM N.A.
 RX MEDLINE=98029496; PubMed=9384671;
 RA Messerle M., Folio M., Nehls M., Eggert H., Boehm T.;
 RT "Dynamic changes in gene expression during in vitro differentiation of
 RT mouse embryonic stem cells.";
 RL Cytokines Cell. Mol. Ther. 1:139-143 (1995).
 CC [7]

RP SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND
 RP MUTAGENESIS OF 1651-ARG--ARG-1654.
 RX MEDLINE=98318619; PubMed=9653148;
 RA Logeat F., Bessia C., Brou C., LeBail O., Jarriault S., Seidah N.G.,
 RA Israel A.;
 RT "The Notch1 receptor is cleaved constitutively by a furin-like
 RT convertase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112 (1998).
 CC [8]
 RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 RT proteolysis.";
 RL J. Biol. Chem. 276:40268-40273 (2001).
 CC [9]
 RP POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031 (2001).
 CC [10]
 RP INTERACTION WITH DTX1 AND DTX2.
 RX MEDLINE=21123790; PubMed=11226752;
 RA Kishi N., Tang Z., Maeda Y., Hirai A., Mo R., Ito M., Suzuki S.,
 RA Nakao K., Kinoshita T., Kadesch T., Hui C.-C., Artavanis-Tsakonas S.,
 RA Okano H., Matsuno K.;
 RT "Murine homologs of deltex define a novel gene family involved in
 RT vertebrate Notch signaling and neurogenesis.";
 RL Int. J. Dev. Neurosci. 19:21-35 (2001).
 CC [-]
 RP FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). May play an essential role in
 CC postimplantation development, probably in some aspect of cell
 CC specification and/or differentiation. May be involved in mesoderm
 CC development, somit formation and neurogenesis. Involved in the
 CC maturation of both CD4+ and CD8+ cells in the thymus.
 CC [-] SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds. Interacts with DTX1 and DTX2.
 CC [-] SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC [-] ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q01705-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q01705-2; Sequence=VSP_001402, VSP_001403, VSP_001404;
 CC Note=No experimental confirmation available;
 CC [-] TISSUE SPECIFICITY: Highly expressed in the brain, lung and
 CC thymus. Expressed at lower levels in the spleen, bone-marrow,
 CC spinal cord, eyes, mammary gland, liver, intestine, skeletal
 CC muscle, kidney and heart.
 CC [-] DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 dpc By
 CC 8.5 dpc highly expressed in presomitic mesoderm, mesenchyme and
 CC endothelial cells, while much lower levels are seen in the
 CC neuroepithelium. Between 9.5-10.5 dpc expressed at high levels in
 CC the neuroepithelium. At 13.5 dpc expressed in the surface
 CC ectoderm, eye and developing whisker follicles.
 CC [-] PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a

notch-derived peptide containing the intracellular domain (NICD)
from the membrane.

-!- PTM: Phosphorylated.
-!- SIMILARITY: Belongs to the NOTCH family.
-!- SIMILARITY: Contains 36 EGF-like domains.
-!- SIMILARITY: Contains 3 Lin/Notch repeats.
-!- SIMILARITY: Contains 5 ANK repeats.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; Z11886; CAA77941.1; -;
EMBL; L02613; AAK14898.1; -;
EMBL; X68278; CAA48339.1; -;
EMBL; AJ238029; CAB40733.1; -;
EMBL; X82562; CAA57909.1; -;
PIR; A46019; A46019;
FIR; B49175; B49175;
HSSP; P00740; 1EDM.
MGD; MGI:97363; Notch1.
GO; GO:0005887; C:integral to plasma membrane; IC.
GO; GO:0005515; P:protein binding; IPI.
GO; GO:0030154; P:cell differentiation; IMP.
GO; GO:0007386; P:compartment specification; IMP.
GO; GO:0007219; P:N signaling pathway; IC.
GO; GO:0045944; P:positive regulation of transcription from P. .; IDA.
InterPro; IPR002110; ANK.
InterPro; IPR000152; Aex_hydroxyl_S.
InterPro; IPR000742; EGF_2.
InterPro; IPR001881; EGF_Ca.
InterPro; IPR001438; EGF_II.
InterPro; IPR006209; EGF_like.
InterPro; IPR002049; Laminin_EGF.
InterPro; IPR008297; Notch.
InterPro; IPR000800; Notch_dom.
Pfam; PF00023; ank; 7.
Pfam; PF00008; EGF; 35.
Pfam; PF00086; notch; 3.
Pfam; PF000279; Notch; 1.
PRINTS; PR00010; EGFLOOD.
PRINTS; PR00011; EGF_LAMININ.
PRINTS; PR01452; NOTCH.
SMART; SM00248; ANK; 6.
SMART; SM00179; EGF_CA; 24.
SMART; SM00004; NL; 2.
PROSITE; PS50297; ANK_REPEAT; 1.
PROSITE; PS50088; ANK_REPEAT; 2.
PROSITE; PS00010; ASX_HYDROXYL; 22.
PROSITE; PS00022; EGF_1; 34.
PROSITE; PS01186; EGF_2; 27.
PROSITE; PS50026; EGF_3; 36.
PROSITE; PS01187; EGF_CA; 21.
Receptor; Transcription regulation; Activator; Differentiation;
Developmental protein; Repeat; ANK repeat; EGF-like domain;
Transmembrane; Glycoprotein; Signal; Phosphorylation;
Alternative splicing.

FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT CHAIN 1711 2531 NOTCH EXTRACELLULAR TRUNCATION.
FT CHAIN 1744 2531 NOTCH INTRACELLULAR DOMAIN.
FT DOMAIN 19 1725 EXTRACELLULAR (POTENTIAL).

Query Match 1.3%; Score 9; DB 1; Length 2531;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 466 PCLHGOTCQ 474
|||||||

Db 1029 PCLHGOTCQ 1037

RESULT 5
NTC1 RAT
ID NTC1_RAT STANDARD; PRT; 2531 AA.
AC Q07008;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1).
GN NOTCH1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Schwann cell;
RX MEDLINE=92111383; PubMed=1764995;
RA Weinmaster G., Roberts V.J., Lemke G.;
RT "A homolog of Drosophila Notch expressed during mammalian
development".
RL Development 113:199-205 (1991).
RN [2]
RP REVISIONS TO 1652-1653.
RA Weinmaster G.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP FUNCTION.
RX MEDLINE=21094508; PubMed=11182080;
RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
RA Honjo T.;
RT "Notchi and Notch3 instructively restrict bFGF-responsive multipotent
neural progenitor cells to an astroglial fate".
RL Neuron 29:45-55 (2001).
RN [4]
RP TISSUE SPECIFICITY.
RX MEDLINE=93202015; PubMed=1295745;
RA Weinmaster G., Roberts V.J., Lemke G.;
RT "Notch2: a second mammalian Notch gene".
RL Development 116:931-941 (1992).
RN [5]
RP TISSUE SPECIFICITY.
RX MEDLINE=21331789; PubMed=11438922;
RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
RT "Expression patterns of Notchi, Notch2, and Notch3 suggest multiple
functional roles for the Notch-DSL signaling system during brain
development".
RL J. Comp. Neurol. 436:167-181 (2001).
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). Acts instructively to control
CC the cell fate determination of CNS multipotent progenitor cells,
CC resulting in astroglial induction and neuron/oligodendrocyte
CC suppression.
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus (By
CC similarity).
CC -!- TISSUE SPECIFICITY: Expressed in the brain, kidney and spleen.
CC , Expressed in postnatal central nervous system (CNS) germinal zones
CC and, in early postnatal life, within numerous cells throughout the
CC CNS. Found in both subventricular and ventricular germinal zones.
CC -!- DEVELOPMENTAL STAGE: In the embryo, highest levels occur between
CC days 12 and 14 and decrease rapidly to much lower levels in the
CC adult.

01-OCT-1996 (Rel. 34, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Sulfate adenylyltransferase subunit 2 (EC 2.7.7.4) (Sulfate adenylyltransferase) (SAR) (ATP-sulfurylase small subunit).
CYSD OR RV1285 OR MT1323 OR MTCY373.04 OR MB1316.
Mycobacterium tuberculosis, and
Mycobacterium bovis.
OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F., Badock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jags K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";
RT Nature 393:537-544(1998).
RL [2]
SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains.";
RT Nature 393:537-544(1998).
RL [3]
SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H., Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S., Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R., Parkhill J., Barrell B.G., Cole S.V., Gordon S.V., Hewinson R.G.;
RA "The complete genome sequence of Mycobacterium bovis.";
RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -1- CATALYTIC ACTIVITY: ATP + sulfate = diphosphate + adenylylsulfate.
CC reaction occurs early in the sulfate activation pathway. This
CC biosynthetic pathway
CC -1- SUBUNIT: Heterodimer composed of cySD, the smaller subunit, and
CC cySN (By similarity).
CC -1- SIMILARITY: Belongs to the PAPS reductase family. Cysd subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z73419; CAA97751.1; ALT INIT.
CC EMBL; AE007007; AAK45584.1; -;
CC EMBL; BX248338; CAD94177.1; ALT_INIT.
CC TIGR; MT1323; -;
CC TubercuList; Rv1285; -;
CC HANAB; MF_00064; -; 1.
CC InterPro; IPR002500; PAPS reduct.
CC Pfam; PF01507; PAPS reduct; 1.
CC Cysteine biosynthesis; Transferase; Nucleotidyltransferase;

KW Complete proteome.
SQ SEQUENCE 309 AA; 34888 MW; 75D965D35F1EC284 CRC64;
Query Match 1.1%; Score 8; DB 1; Length 309;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 279 VLLFSGGK 286
DB 41 VLLFSGGK 48
RESULT 7
AMPA_BACSU STANDARD; PRT; 500 AA.
ID AMPA_BACSU
AC 032106;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable cytosol aminopeptidase (EC 3.4.11.1) (Leucine aminopeptidase)
DE (LAP) (Leucyl aminopeptidase).
DE PEPA OR BSU32050.
GN Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogaewara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bartero M.G., Bessieres P., Bolotin A., Borchert S., Borriess R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T., Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Guiseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B., Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpatra P., Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A., Viari A., Wambutt R., Wedler H., Wedler H., Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.P., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.";
RL Nature 390:249-256(1997).
CC -1- FUNCTION: Presumably involved in the processing and regular turnover of intracellular proteins. Catalyzes the removal of unsubstituted N-terminal amino acids from various peptides (By similarity).
CC -1- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-|-Xbb-, in which xaa is preferably Leu, but may be other amino acids including Pro although not Arg or Lys, and Xbb may be Pro.
CC -1- COFACTOR: Binds 2 manganese ions per subunit (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to peptidase family M17.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL; Z99120; CAB15195.1; --
 DR PIR; F70012; F70012.
 DR HSP; P00727; ILAM.
 DR MEROPS; M17.UFW; --
 DR Subtilist; BG13970; pepA.
 DR HAMAP; MF_00181; --; 1.
 DR InterPro; IPR000819; Peptidase M17 C.
 DR InterPro; IPR008283; Peptidase M17 N.
 DR Pfam; PF00883; Peptidase M17; 1.
 DR Pfam; PF02789; Peptidase M17 N; 1.
 DR PRINTS; PR00481; LAMNOPPTDASE.
 DR PROSITE; PS00631; CYTOSOL_AP; 1.
 KW Hydrolase; Aminopeptidase; Manganese; Complete proteome.
 FT METAL 261
 FT METAL 266
 FT METAL 266
 FT METAL 284
 FT METAL 343
 FT METAL 345
 FT METAL 345
 FT ACT_SITE 273
 FT ACT_SITE 347
 FT ACT_SITE 347
 SQ SEQUENCE 500 AA; 53657 MW; 3E82968F6656559 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 500;
 Best Local Similarity 100.0%; Pred.No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 653 DVATLTGG 660

DB 367 DVATLTGG 374

RESULT 8

PCGN RAT STANDARD; PRT; 1257 AA.
 AC P55067;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3)
 DE (245 kDa early postnatal core glycoprotein) [Contains: 150 kDa adult
 DE core glycoprotein].
 DE CSPG3 OR NCAN.
 GN Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-Sprague-Dawley; TISSUE-Brain;
 RX MEDLINE=92406907; PubMed=1326557;
 RA Rauch U., Karthikeyan L., Maurel P., Margolis R.U., Margolis R.K.;
 RT "Cloning and primary structure of neurocan, a developmentally
 RT regulated, aggregating chondroitin sulfate proteoglycan of brain."
 RL J Biol. Chem. 267:19536-19547(1992).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=94230574; PubMed=7513709;
 RA Friedlander D.R., Milev P., Karthikeyan L., Margolis R.K.,
 RA Margolis R.U., Grumet M.;
 RT "The neuronal chondroitin sulfate proteoglycan neurocan binds to the
 RT neural cell adhesion molecules Ng-CAM/11/NILE and N-CAM, and inhibits
 RT neuronal adhesion and neurite outgrowth."
 RL J. Cell Biol. 125:669-680(1994).
 CC -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
 CC development by binding to neural cell adhesion molecules (NG-CAM
 CC and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
 CC acid.

CC -!- TISSUE SPECIFICITY: Early postnatal and adult brain; not expressed
 CC in kidney, lung, liver and muscle.
 CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND
 CC O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
 CC -!- PTM: TWO ISOFORMS WERE FOUND THAT PROBABLY ARISE BY PROTEOLYTIC
 CC DEGRADATION. THE LARGE ISOFORM IS PREDOMINANT IN EARLY POSTNATAL
 CC BRAIN, THE SMALL ISOFORM IS FOUND IN ADULT BRAIN.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 2 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M97161; AAC37679.1; --
 DR PIR; S28764; S28764.
 DR HSP; P00740; LEDM.
 DR InterPro; IPR002353; AntifreezeII.
 DR InterPro; IPR000152; Asx hydroxyl_s.
 DR InterPro; IPR000742; EGF_2
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF-Like.
 DR InterPro; IPR007110; IG-Like.
 DR InterPro; IPR003599; IG.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00047; IG; 1.
 DR Pfam; PF00059; lectin_c; 1.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 2.
 DR PRINTS; PR00356; ANTIFREEZEII.
 DR ProDom; PD000918; Link; 2.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CUECT; 1.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 2.
 DR PROSITE; PS00010; ASX HYDROXYL; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
 DR PROSITE; PS00041; C-TYPE LECTIN_2; 1.
 DR PROSITE; PS00022; EGF_1; 3.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS00026; EGF_3; 2.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS00835; IG LINK; 1.
 DR PROSITE; PS01241; LINK; 2.
 KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
 KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 1257 NEUROCAN CORE PROTEIN.
 FT CHAIN 639 1257 150 kDa ADULT CORE GLYCOPROTEIN.
 FT DOMAIN 37 157 IG-LIKE V-TYPE.
 FT DOMAIN 158 253 LINK 1.
 FT DOMAIN 259 355 LINK 2.
 FT DOMAIN 949 985 EGF-LIKE 1.
 FT DOMAIN 987 1023 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1025 1154 C-TYPE LECTIN.
 FT DOMAIN 1155 1213 SUSHI.
 FT DOMAIN 58 139 BY SIMILARITY.
 FT DISULFID 181 252 BY SIMILARITY.
 FT DISULFID 205 226 BY SIMILARITY.
 FT DISULFID 279 354 BY SIMILARITY.

```

FT DISULFID 303 324 BY SIMILARITY.
FT DISULFID 933 964 BY SIMILARITY.
FT DISULFID 938 973 BY SIMILARITY.
FT DISULFID 975 984 BY SIMILARITY.
FT DISULFID 1029 1040 BY SIMILARITY.
FT DISULFID 1057 1149 BY SIMILARITY.
FT DISULFID 1125 1141 BY SIMILARITY.
FT DISULFID 1156 1199 BY SIMILARITY.
FT DISULFID 1185 1212 BY SIMILARITY.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 944 944 O-LINKED (XYL. . .) (CHONDROITIN SULFATE).
FT CARBOHYD 967 967 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1164 1164 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1257 AA; 135544 MW; 992B33DCFA19E1B CRC64;

Query Match 1.1%; Score 8; DB 1; Length 1257;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGGTC 473
DB 957 PCLHGGTC 964

RESULT 9
PCGN MOUSE
ID PCGN MOUSE STANDARD; PRT; 1268 AA.
AC P55066;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3).
GN CSPP3 OR NCAN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Brain;
RX MEDLINE=96039250; PubMed=7490074;
RA Rauch U., Grimpe B., Kulbe G., Arnold-Ammer I., Beier D.,
RA Passler R.;
RT "Structure and chromosomal localization of the mouse neurocan gene.";
RL Genomics 28:405-410(1995).
CC -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
CC development by binding to neural cell adhesion molecules (NG-CAM
CC and N-CAM). Chondroitin sulfate proteoglycan, binds to hyaluronic
CC acid.
CC -!- TISSUE SPECIFICITY: Brain.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X84727; CAA59216.1; -
CC PIR; S52781; S52781.
CC HSSP; P00740; 1EDM.
CC MGD; MGI:104694; Cspg3.
CC InterPro; IPR002353; AntifreezeZell.
CC InterPro; IPR000152; Asx_hydroxyl_S.

```

```

DR InterPro; IPR000742; EGF 2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR007110; Ig-Like.
DR InterPro; IPR003599; Ig-Like.
DR InterPro; IPR001394; Lentin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi_1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PRO0356; ANTIFREEZEII.
DR PRINTS; PRO1265; LINKMODULE.
DR PRODOM; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00409; IG_1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.
DR PROSITE; PS00041; C-TYPE LECTIN 2; 1.
DR PROSITE; PS00022; EGF_1_3.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00026; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW EGF-like domain; Calcium; Repeat; Lentin; Sushi; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1268 NEUROCAN CORE PROTEIN.
FT DOMAIN 37 157 IG-LIKE V-TYPE.
FT DOMAIN 158 253 LINK 1.
FT DOMAIN 259 355 LINK 2.
FT DOMAIN 960 996 EGF-LIKE 1.
FT DOMAIN 998 1034 EGF-LIKE 2.
FT DOMAIN 1036 1165 C-TYPE LECTIN.
FT DOMAIN 1166 1224 SUSHI.
FT DISULFID 58 139 BY SIMILARITY.
FT DISULFID 181 252 BY SIMILARITY.
FT DISULFID 205 226 BY SIMILARITY.
FT DISULFID 279 354 BY SIMILARITY.
FT DISULFID 303 324 BY SIMILARITY.
FT DISULFID 964 975 BY SIMILARITY.
FT DISULFID 969 984 BY SIMILARITY.
FT DISULFID 986 995 BY SIMILARITY.
FT DISULFID 1040 1051 BY SIMILARITY.
FT DISULFID 1068 1160 BY SIMILARITY.
FT DISULFID 1136 1152 BY SIMILARITY.
FT DISULFID 1167 1210 BY SIMILARITY.
FT DISULFID 1196 1223 BY SIMILARITY.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 742 742 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 978 978 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1175 1175 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1268 AA; 137200 MW; 3014E8E20A2FAEC CRC64;

Query Match 1.1%; Score 8; DB 1; Length 1268;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGGTC 473
DB 968 PCLHGGTC 975

RESULT 10
PCGN HUMAN
ID PCGN HUMAN STANDARD; PRT; 1321 AA.

```


AC O14594; Q9UPK6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Neucan core protein precursor (Chondroitin sulfate proteoglycan 3).
 GN CSPG3 OR NCAN OR NEUR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99013874; PubMed=9795216;
 RA Prange C.K., Pennacchio L.A., Lieuallen K., Fan W., Lennon G.G.;
 RT "Characterization of the human neucan gene, CSPG3.";
 RL Gene 221:199-205(1998).
 RN [2]
 RP SEQUENCE OF 1-990 AND 1007-1321 FROM N.A., AND VARIANT ALA-1254.
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
 RA Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Gaines J.,
 RA Dangnan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
 RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
 RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
 RA Krommiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
 RA Kobayashi A., Olsen A.S., Carrano A.V.;
 RT "Sequence analysis of an -1 Mb region containing the MEF2B gene in
 19p12.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBSJ databases.
 CC -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
 development by binding to neural cell adhesion molecules (NG-CAM
 and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
 acid.
 CC -!- TISSUE SPECIFICITY: Brain.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 2 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 link domains.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF036547; AAC80576.1; -;
 CC EMBL; AC003110; AAB86655.1; -;
 CC EMBL; AC005254; AAC25581.1; -;
 CC HSP; P00740; 1EDM.
 CC Genew; HGNC:2465; CSPG3.
 CC MIM; 600826; -;
 CC InterPro; IPR000152; Asx hydroxyl_s.
 CC InterPro; IPR000742; EGF 2.
 CC InterPro; IPR001881; EGF Ca.
 CC InterPro; IPR006209; EGF-like.
 CC InterPro; IPR007110; Ig-like.
 CC InterPro; IPR003599; Ig.
 CC InterPro; IPR001304; Lectin_C.
 CC InterPro; IPR005338; Link.
 CC InterPro; IPR000436; Sushi_SCR_CCP.
 CC Pfam; PF00008; EGF; 2.
 CC Pfam; PF00047; Ig; 1.
 CC Pfam; PF00059; lectin_c; 1.
 CC Pfam; PF00084; sushi; 1.
 CC Pfam; PF00193; Xlink; 2.
 CC PRINTS; PR01265; LINKMODULE.
 CC ProDom; PD000918; Link; 2.
 CC SMART; SM00032; CCP; 1.
 CC SMART; SM00034; CLECT; 1.

DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 2.
 DR PROSITE; PS00010; ASX HYDROXYL; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.
 DR PROSITE; PS00041; C-TYPE LECTIN_2; 1.
 DR PROSITE; PS00022; EGF_1; 3.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS00026; EGF_3; 2.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS00835; IG-LIKE; 1.
 DR PROSITE; PS01241; LINK; 2.
 KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
 KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal; Polymorphism.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 1321 NEUCAN CORE PROTEIN.
 FT DOMAIN 38 153 IG-LIKE V-TYPE.
 FT DOMAIN 159 254 LINK 1.
 FT DOMAIN 260 356 LINK 2.
 FT DOMAIN 1008 1044 EGF-LIKE 1.
 FT DOMAIN 1046 1082 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1084 1213 C-TYPE LECTIN.
 FT DOMAIN 1214 1272 SUSHI.
 FT DISULFID 59 140 BY SIMILARITY.
 FT DISULFID 182 253 BY SIMILARITY.
 FT DISULFID 206 227 BY SIMILARITY.
 FT DISULFID 280 325 BY SIMILARITY.
 FT DISULFID 304 325 BY SIMILARITY.
 FT DISULFID 1012 1023 BY SIMILARITY.
 FT DISULFID 1017 1032 BY SIMILARITY.
 FT DISULFID 1034 1043 BY SIMILARITY.
 FT DISULFID 1088 1099 BY SIMILARITY.
 FT DISULFID 1116 1208 BY SIMILARITY.
 FT DISULFID 1184 1200 BY SIMILARITY.
 FT DISULFID 1215 1258 BY SIMILARITY.
 FT DISULFID 1244 1271 BY SIMILARITY.
 FT CARBOHYD 122 122 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1026 1026 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1223 1223 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARIANT 1254 1254 V -> A (in dbSNP:1064389).
 FT CONFLICT 1234 1234 N -> Y (IN REF. 2).
 FT CONFLICT 1282 1282 G -> R (IN REF. 2).
 SQ SEQUENCE 1321 AA; 142972 MW; 2EF47F823DB980B8 CRC64;
 Query Match 1.1%; Score 8; DB 1; Length 1321;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 466 PCLHGTC 473
 DB 1016 PCLHGTC 1023
 RESULT 11
 AGRI_DISOM
 ID AGRI_DISOM STANDARD; PRT; 1328 AA.
 AC Q90404;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Agrin (Fragment).
 GN AGRN
 OS Discoptye ornata (Electric ray).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Squalia; Hymnosqualia; Pristioraja; Batoida;
 OC Torpediniformes; Narcinoidae; Narcinidae; Discopyge.
 OX NCBI_TaxID=7785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smith M.A., Magill-Solc C., Rupp F., Yao Y.-M.M., Schilling J.W.,
 RA Snow P., McMahon U.J.;

RT "Isolation and characterization of a cDNA that encodes an agrin homolog in the marine ray.";
 RL Mol. Cell. Neurosci. 3:406-417(1992).
 CC -!- FUNCTION: Component of the basal lamina that causes the aggregation of acetylcholine receptors and acetylcholine-esterase on the surface of muscle fibers of the neuromuscular junction (By similarity).
 CC -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular junction (By similarity).
 CC -!- SIMILARITY: Contains at least 2 laminin EGF-like domains.
 CC -!- SIMILARITY: Contains at least 3 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 SEA domain.
 CC -!- SIMILARITY: Contains 3 laminin G-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L01423; AAA49224.1; -
 DR PIR; T43060; T43060.
 DR HSSP; P00740; 1EDM.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR008985; ConA_like_lec_gl.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR008209; EGF_like.
 DR InterPro; IPR003645; FOLN.
 DR InterPro; IPR002350; FOLN.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR000082; SEA_domain.
 DR Pfam; PF00008; EGF_4.
 DR Pfam; PF00050; kazal; 2.
 DR Pfam; PF00053; laminin_EGF; 2.
 DR Pfam; PF00054; laminin_G; 3.
 DR Pfam; PF01390; SEA; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR SMART; SM00180; EGF_Lam; 2.
 DR SMART; SM00274; FOLN; 2.
 DR SMART; SM00280; KAZAL; 2.
 DR SMART; SM00282; LamG; 3.
 DR SMART; SM00200; SEA; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00022; EGF_1; 5.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50026; EGF_3; 4.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
 DR PROSITE; PS50024; SEA; 1.
 DR Glycoprotein; EGF-like domain; Repeat; Laminin EGF-like domain.
 KW NON_TER 1 1
 FT DOMAIN 79 132 LAMININ EGF-LIKE 1.
 FT DOMAIN 133 179 LAMININ EGF-LIKE 2.
 FT DOMAIN 411 533 SEA.
 FT DOMAIN 608 644 EGF-LIKE 1.
 FT DOMAIN 649 825 LAMININ G-LIKE 1.
 FT DOMAIN 865 902 EGF-LIKE 2.
 FT DOMAIN 914 1096 LAMININ G-LIKE 2.
 FT DOMAIN 1097 1135 EGF-LIKE 3.
 FT DOMAIN 1146 1324 LAMININ G-LIKE 3.
 FT DISULFID 79 91 BY SIMILARITY.
 FT DISULFID 81 98 BY SIMILARITY.
 FT DISULFID 100 109 BY SIMILARITY.
 FT DISULFID 112 130 BY SIMILARITY.
 FT DISULFID 133 145 BY SIMILARITY.
 FT DISULFID 135 152 BY SIMILARITY.
 FT DISULFID 154 163 BY SIMILARITY.
 FT DISULFID 166 177 BY SIMILARITY.
 FT DISULFID 612 623 BY SIMILARITY.
 FT DISULFID 617 632 BY SIMILARITY.

FT DISULFID 634 643 BY SIMILARITY.
 FT DISULFID 869 880 BY SIMILARITY.
 FT DISULFID 874 890 BY SIMILARITY.
 FT DISULFID 892 901 BY SIMILARITY.
 FT DISULFID 1101 1114 BY SIMILARITY.
 FT DISULFID 1108 1123 BY SIMILARITY.
 FT DISULFID 1125 1134 BY SIMILARITY.
 FT CARBOHYD 63 63 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 168 168 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 279 279 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 369 369 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1328 AA; 144018 MW; 79D81C1AF2A71C18 CRC64;
 Query Match 1.1%; Score 8; DB 1; Length 1328;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 466 PCLHGTC 473
 DB 616 PCLHGTC 623
 RESULT 12
 AGRI_CHICK
 ID AGRI_CHICK STANDARD; PRT; 1955 AA.
 AC P31696;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DR AGRI precursor.
 GN AGRI.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]_TaxID=9031;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=92232297; PubMed=1314620;
 RA Tsim K.W.K., Ruegg M.A., Escher G., Kroeger S., McMahon U.J.;
 RT "cDNA that encodes active agrin.";
 RL Neuron 8:677-689(1992).
 RN [2]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=92232298; PubMed=1314621;
 RA Ruegg M.A., Tsim K.W.K., Horton S.E., Kroeger S., Escher G.,
 RA Gensch E.M., McMahon U.J.;
 RT "The agrin gene codes for a family of basal lamina proteins that differ in function and distribution.";
 RL Neuron 8:691-699(1992).
 CC -!- FUNCTION: Component of the basal lamina that causes the aggregation of acetylcholine receptors and acetylcholine-esterase on the surface of muscle fibers of the neuromuscular junction.
 CC -!- SUBUNIT: Binds to laminin.
 CC -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular junction.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Comment=Additional isoforms seem to exist. Isoforms differ in their acetylcholine receptor clustering activity;
 CC Name=1;
 CC IsoId=P31696-1; Sequence=Displayed;
 CC Name=2; Synonyms=Agri-related protein 1;
 CC IsoId=P31696-2; Sequence=VSP_001370;
 CC Name=3; Synonyms=Agri-related protein 2;
 CC IsoId=P31696-3; Sequence=VSP_001369, VSP_001370;
 CC -!- PTM: Contains heparan sulfate chains as well as N-linked and O-linked oligosaccharides (By similarity).
 CC -!- SIMILARITY: Contains 9 kazal-like domains.
 CC -!- SIMILARITY: Contains 2 laminin EGF-like domains.
 CC -!- SIMILARITY: Contains 4 EGF-like domains.


```

RX MEDLINE=902631104; PubMed=2344615;
RA Tepass U., Theres C., Knust E.;
RT "Crumbs encodes an EGF-like protein expressed on apical membranes of
RT Drosophila epithelial cells and required for organization of
RT epithelia.";
RL Cell 61:787-799(1990).
RN [2]
RP SEQUENCE OF 1663-1955 FROM N.A.
RC TISSUE=Embryo;
RA MEDLINE=87218537; PubMed=3107986;
RA Knust E., Dietrich U., Tepass U., Bremer K.A., Weigal D.,
RA Vaessin H., Campos-Ortega J.A.;
RT "EGF homologous sequences encoded in the genome of Drosophila
RT melanogaster, and their relation to neurogenic genes.";
RL EMBO J. 6:761-766(1987).
CC -1- FUNCTION: May play a role in the development of epithelia,
CC possibly for the establishment and/or maintenance of cell
CC polarity. It may act as a signal.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
CC -1- SIMILARITY: Contains 29 EGF-like domains.
CC -1- SIMILARITY: Contains 3 laminin G-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M33753; AAA28428.1; ALT_SEQ.
DR EMBL; X05144; CAA28793.1; -.
DR PIR; A35672; A35672.
DR PIR; B26637; B26637.
DR HSP; P00740; IEDM.
DR FlyBase; FBgn0000368; crb.
DR GO; GO:0016324; Crapical plasma membrane; NAS.
DR GO; GO:0016327; Crapicolateral plasma membrane; IDA.
DR GO; GO:0007163; Prestablishment and/or maintenance of cell po. .; IMP.
DR GO; GO:0016332; Prestablishment and/or maintenance of polarit. .; IMP.
DR GO; GO:0016334; Prestablishment and/or maintenance of polarit. .; IMP.
DR GO; GO:0045494; Pphotoreceptor maintenance; IMP.
DR GO; GO:0042052; Prhodanore development; NAS.
DR GO; GO:0045186; Pizonula adherens assembly; IMP.
DR GO; GO:0045218; Pizonula adherens maintenance; IMP.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_Ii.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00008; EGF; 26.
DR Pfam; PF00054; laminin_G; 3.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR00011; EGFLAMININ.
DR SMART; SM00179; EGF_CA; 11.
DR SMART; SM00282; LamG; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 15.
DR PROSITE; PS00022; EGF_1; 26.
DR PROSITE; PS01186; EGF_2; 17.
DR PROSITE; PS00026; EGF_3; 27.
DR PROSITE; PS01187; EGF_CA; 12.
DR PROSITE; PS00025; LamG_DOMAIN; 3.
KW Differentiation; Repeat; EGF-like domain; Transmembrane;
KW Glycoprotein; Signal; Phosphorylation.
FT SIGNAL 1 90
FT CHAIN 91 2139 CRUMBS PROTEIN.
FT DOMAIN 91 2084 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2085 2111 POTENTIAL.
FT DOMAIN 2112 2139 CYTOPLASMIC (POTENTIAL).
FT -----
FT DOMAIN 267 303 EGF-LIKE 1.
FT DOMAIN 306 343 EGF-LIKE 2.
FT DOMAIN 348 386 EGF-LIKE 3.
FT DOMAIN 388 425 EGF-LIKE 4. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 427 463 EGF-LIKE 5.
FT DOMAIN 464 500 EGF-LIKE 6.
FT DOMAIN 501 532 EGF-LIKE 7.
FT DOMAIN 545 581 EGF-LIKE 8.
FT DOMAIN 582 611 EGF-LIKE 9.
FT DOMAIN 646 682 EGF-LIKE 10.
FT DOMAIN 687 723 EGF-LIKE 11. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 725 761 EGF-LIKE 12. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 763 800 EGF-LIKE 13. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 802 838 EGF-LIKE 14. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 840 902 EGF-LIKE 15.
FT DOMAIN 904 940 EGF-LIKE 16. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 942 978 EGF-LIKE 17. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 980 1021 EGF-LIKE 18. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1023 1205 LAMININ G-LIKE 1.
FT DOMAIN 1207 1243 EGF-LIKE 19.
FT DOMAIN 1250 1480 LAMININ G-LIKE 2.
FT DOMAIN 1481 1517 EGF-LIKE 20.
FT DOMAIN 1518 1558 EGF-LIKE 21.
FT DOMAIN 1559 1758 LAMININ G-LIKE 3.
FT DOMAIN 1759 1795 EGF-LIKE 22.
FT DOMAIN 1833 1871 EGF-LIKE 23. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1874 1915 EGF-LIKE 24. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1915 1951 EGF-LIKE 25.
FT DOMAIN 1951 1989 EGF-LIKE 26.
FT DOMAIN 1989 2029 EGF-LIKE 27. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2030 2070 EGF-LIKE 28. CALCIUM-BINDING (POTENTIAL).
FT DISULFID 271 282 BY SIMILARITY.
FT DISULFID 276 291 BY SIMILARITY.
FT DISULFID 293 302 BY SIMILARITY.
FT DISULFID 310 321 BY SIMILARITY.
FT DISULFID 315 331 BY SIMILARITY.
FT DISULFID 333 342 BY SIMILARITY.
FT DISULFID 352 363 BY SIMILARITY.
FT DISULFID 376 385 BY SIMILARITY.
FT DISULFID 392 403 BY SIMILARITY.
FT DISULFID 397 412 BY SIMILARITY.
FT DISULFID 414 424 BY SIMILARITY.
FT DISULFID 431 442 BY SIMILARITY.
FT DISULFID 436 451 BY SIMILARITY.
FT DISULFID 453 462 BY SIMILARITY.
FT DISULFID 468 479 BY SIMILARITY.
FT DISULFID 473 488 BY SIMILARITY.
FT DISULFID 490 499 BY SIMILARITY.
FT DISULFID 505 515 BY SIMILARITY.
FT DISULFID 509 520 BY SIMILARITY.
FT DISULFID 522 531 BY SIMILARITY.
FT DISULFID 549 562 BY SIMILARITY.
FT DISULFID 556 569 BY SIMILARITY.
FT DISULFID 571 580 BY SIMILARITY.
FT DISULFID 586 597 BY SIMILARITY.
FT DISULFID 591 602 BY SIMILARITY.
FT DISULFID 604 610 BY SIMILARITY.
FT DISULFID 613 624 BY SIMILARITY.
FT DISULFID 618 634 BY SIMILARITY.
FT DISULFID 636 645 BY SIMILARITY.
FT DISULFID 652 664 BY SIMILARITY.
FT DISULFID 659 673 BY SIMILARITY.
FT DISULFID 675 684 BY SIMILARITY.
FT DISULFID 691 702 BY SIMILARITY.
FT DISULFID 696 711 BY SIMILARITY.
FT DISULFID 713 722 BY SIMILARITY.
FT DISULFID 729 740 BY SIMILARITY.
FT DISULFID 734 749 BY SIMILARITY.
FT DISULFID 751 760 BY SIMILARITY.
FT DISULFID 767 778 BY SIMILARITY.
FT DISULFID 772 787 BY SIMILARITY.

```

```

FT DISULFID 789 799 BY SIMILARITY.
FT DISULFID 806 817 BY SIMILARITY.
FT DISULFID 811 826 BY SIMILARITY.
FT DISULFID 828 837 BY SIMILARITY.
FT DISULFID 844 855 BY SIMILARITY.
FT DISULFID 849 890 BY SIMILARITY.
FT DISULFID 892 901 BY SIMILARITY.
FT DISULFID 908 919 BY SIMILARITY.
FT DISULFID 913 928 BY SIMILARITY.
FT DISULFID 930 939 BY SIMILARITY.
FT DISULFID 946 957 BY SIMILARITY.
FT DISULFID 952 966 BY SIMILARITY.
FT DISULFID 968 977 BY SIMILARITY.
FT DISULFID 984 995 BY SIMILARITY.
FT DISULFID 999 1009 BY SIMILARITY.
FT DISULFID 1011 1020 BY SIMILARITY.
FT DISULFID 1211 1222 BY SIMILARITY.
FT DISULFID 1216 1231 BY SIMILARITY.
FT DISULFID 1233 1242 BY SIMILARITY.
FT DISULFID 1485 1496 BY SIMILARITY.
FT DISULFID 1490 1505 BY SIMILARITY.
FT DISULFID 1507 1516 BY SIMILARITY.
FT DISULFID 1763 1774 BY SIMILARITY.
FT DISULFID 1768 1783 BY SIMILARITY.
FT DISULFID 1785 1794 BY SIMILARITY.
FT DISULFID 1801 1812 BY SIMILARITY.
FT DISULFID 1806 1821 BY SIMILARITY.
FT DISULFID 1823 1832 BY SIMILARITY.
FT DISULFID 1839 1850 BY SIMILARITY.
FT DISULFID 1844 1859 BY SIMILARITY.
FT DISULFID 1861 1870 BY SIMILARITY.
FT DISULFID 1878 1889 BY SIMILARITY.
FT DISULFID 1883 1903 BY SIMILARITY.
FT DISULFID 1905 1914 BY SIMILARITY.
FT DISULFID 1919 1930 BY SIMILARITY.
FT DISULFID 1924 1939 BY SIMILARITY.
FT DISULFID 1941 1950 BY SIMILARITY.
FT DISULFID 1957 1968 BY SIMILARITY.
FT DISULFID 1962 1977 BY SIMILARITY.

Query Match 1.1%; Score 8; DB 1;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 PCQNGQC 173
Db 1661 PCQNGQC 1668

RESULT 14
ID NTC3 HUMAN STANDARD; PRT; 2321 AA.
AC Q9UM47; Q9UEB3; Q9UEJ3; Q9Y6L6;
DI 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
GN NOTCH3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX MEDLINE=97032728; PubMed=8878478;
RA Joutel A., Corpechot C., Ducros A., Vahedi K., Chabriat H., Mouton P.,
RA Alamiwitch S., Domenga V., Cecillion M., Marechal E., Maciazek J.,
RA Vaysiere C., Cruaud C., Cabanis E.-A., Ruchoux M.M., Weissenbach J.,
RA Bach J.-F., Bousser M.-G., Tournier-Lasserre E.;
RT "Notch3 mutations in CADASIL, a hereditary adult-onset condition
causing stroke and dementia.";
RL Nature 383:707-710(1996).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RA Gunel M., Artavanis-Tsakonas S.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCreedy P.M., Skowroneki E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA Phan H., Velasco N., Ganes J., Danganan L., Poundstone P.,
RA Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,
RA Trankheim M., Amico-Keller G., Cosfield J., Duarte S., Lucas S.,
RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,
RA Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,
RA Carrano A.V.;
RT "Sequence analysis of an 1.5 Mb olfactory receptor (OLFR) cluster in
19p13.1";
RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP VARIANTS CADASIL TYR-49; CYS-71; CYS-90; CYS-110; CYS-133; CYS-141;
RP ARG-146; CYS-153; CYS-169; CYS-173; CYS-182; ARG-185; SER-212;
RP GLY-222; TYR-224; CYS-258; TYR-542; CYS-558; CYS-578; CYS-728;
RP CYS-985; CYS-1006; CYS-1031; CYS-1231 AND ARG-1261, AND VARIANTS
RP ARG-170; LEU-496; GLN-1133; MET-1183 AND ALA-2223.
RX MEDLINE=98049753; PubMed=9388399;
RA Joutel A., Vahedi K., Corpechot C., Troesch A., Chabriat H.,
RA Vaysiere C., Cruaud C., Maciazek J., Weissenbach J., Bousser M.-G.,
RA Bach J.-F., Tournier-Lasserre E.;
RT "Strong clustering and stereotyped nature of Notch3 mutations in
CADASIL patients.";
RL Lancet 350:1511-1515(1997).
RN [5]
RP VARIANT CADASIL 114-GLY--PRO-120 DEL.
RX MEDLINE=20264473; PubMed=10802807;
RA Joutel A., Chabriat H., Vahedi K., Domenga V., Vaysiere C.,
RA Ruchoux M.M., Lucas C., Leys D., Bousser M.-G., Tournier-Lasserre E.;
RT "Splice site mutation causing a seven amino acid Notch3 in-frame
deletion in CADASIL.";
RL Neurology 54:1874-1875(2000).
RN [6]
RP IDENTIFICATION OF LIGANDS.
RX MEDLINE=99180765; PubMed=10079256;
RA Grays G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
RT "Human ligands of the Notch receptor.";
RL Am. J. Pathol. 154:785-794(1999).
CC -I- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity).
CC -I- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds (By similarity).
CC -I- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -I- TISSUE SPECIFICITY: Ubiquitously expressed in fetal and adult
CC tissues.
CC -I- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -I- PTM: Phosphorylated (By similarity).
CC -I- DISBASE: Defects in NOTCH3 are associated with cerebral autosomal
CC dominant arteriopathy with subcortical infarcts and
CC leukoencephalopathy (CADASIL) [MIM:125310]. CADASIL causes a type
CC

```

of stroke and dementia of which key features include recurrent
 subcortical ischemic events and vascular dementia.
 -1- SIMILARITY: Belongs to the NOTCH family.
 -1- SIMILARITY: Contains 34 EGF-like domains.
 -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 -1- SIMILARITY: Contains 5 ANK repeats.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; U97669; AAB91371.1; -
 EMBL; AF058900; AAC14346.1; -
 EMBL; AF058881; AAC14346.1; JOINED.
 EMBL; AF058882; AAC14346.1; JOINED.
 EMBL; AF058883; AAC14346.1; JOINED.
 EMBL; AF058884; AAC14346.1; JOINED.
 EMBL; AF058885; AAC14346.1; JOINED.
 EMBL; AF058886; AAC14346.1; JOINED.
 EMBL; AF058887; AAC14346.1; JOINED.
 EMBL; AF058888; AAC14346.1; JOINED.
 EMBL; AF058889; AAC14346.1; JOINED.
 EMBL; AF058890; AAC14346.1; JOINED.
 EMBL; AF058891; AAC14346.1; JOINED.
 EMBL; AF058892; AAC14346.1; JOINED.
 EMBL; AF058893; AAC14346.1; JOINED.
 EMBL; AF058894; AAC14346.1; JOINED.
 EMBL; AF058895; AAC14346.1; JOINED.
 EMBL; AF058896; AAC14346.1; JOINED.
 EMBL; AF058897; AAC14346.1; JOINED.
 EMBL; AF058898; AAC14346.1; JOINED.
 EMBL; AF058899; AAC14346.1; JOINED.
 EMBL; AC004257; AAC04897.1; -
 EMBL; AC004663; AAC15789.1; ALT_INIT.
 PIR; S78549; S78549.
 HSP; P00740; IEDM.
 Genew; HGNC:7893; NOTCH3.
 MIM; 600276; -
 MIM; 125310; -
 InterPro; IPR002110; ANK.
 InterPro; IPR000152; Asx_hydroxyl_S.
 InterPro; IPR000742; EGF_2.
 InterPro; IPR001881; EGF_Ca.
 InterPro; IPR001438; EGF_II.
 InterPro; IPR006209; EGF_Like.
 InterPro; IPR002049; Laminin_EGF.
 InterPro; IPR008297; Notch.
 InterPro; IPR008000; Notch_dom.
 Pfam; PF00023; ank; 6.
 Pfam; PF00008; EGF; 34.
 Pfam; PF00066; notch; 3.
 PIRSF; PIRSF002279; Notch; 1.
 PRINTS; PR00010; EGFLOOD.
 PRINTS; PR00011; EGFAMININ.
 PRINTS; PR01452; NOTCH.
 SMART; SM00248; ANK; 6.
 SMART; SM00179; EGF_Ca; 19.
 SMART; SM00004; NL; 3.
 PROSITE; PS00297; ANK_REPEAT; 1.
 PROSITE; PS00088; ANK_REPEAT; 4.
 PROSITE; PS00010; ASX_HYDROXYL; 18.
 PROSITE; PS00022; EGF_1; 33.
 PROSITE; PS01186; EGF_2; 25.
 PROSITE; PS00026; EGF_3; 34.
 PROSITE; PS01187; EGF_Ca; 16.
 Receptor; Transcription regulation; Activator; Differentiation;
 Developmental protein; Repeat; ANK repeat; EGF-like domain;
 Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 Disease mutation.

FT	SIGNAL	1	39	POTENTIAL.
FT	CHAIN	40	2321	NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
FT	CHAIN	1629	2321	NOTCH EXTRACELLULAR TRUNCATION (BY
FT				SIMILARITY).
FT	CHAIN	1662	2321	NOTCH INTRACELLULAR DOMAIN (BY
FT				SIMILARITY).
FT	DOMAIN	40	1643	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1644	1664	POTENTIAL.
FT	DOMAIN	1665	2321	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	78	118	EGF-LIKE 1.
FT	DOMAIN	119	156	EGF-LIKE 2.
FT	DOMAIN	158	195	EGF-LIKE 3.
FT	DOMAIN	197	234	EGF-LIKE 4.
FT	DOMAIN	236	272	EGF-LIKE 5.
FT	DOMAIN	274	312	EGF-LIKE 6.
FT	DOMAIN	314	350	EGF-LIKE 7.
FT	DOMAIN	351	389	EGF-LIKE 8.
FT	DOMAIN	391	429	EGF-LIKE 9.
FT	DOMAIN	431	467	EGF-LIKE 10.
FT	DOMAIN	469	505	EGF-LIKE 11.
FT	DOMAIN	507	543	EGF-LIKE 12.
FT	DOMAIN	545	580	EGF-LIKE 13.
FT	DOMAIN	582	618	EGF-LIKE 14.
FT	DOMAIN	620	655	EGF-LIKE 15.
FT	DOMAIN	657	693	EGF-LIKE 16.
FT	DOMAIN	695	730	EGF-LIKE 17.
FT	DOMAIN	734	770	EGF-LIKE 18.
FT	DOMAIN	771	808	EGF-LIKE 19.
FT	DOMAIN	810	847	EGF-LIKE 20.
FT	DOMAIN	849	885	EGF-LIKE 21.
FT	DOMAIN	887	922	EGF-LIKE 22.
FT	DOMAIN			EGF-LIKE 23.

Query Match 1.1%; Score 8; DB 1; Length 2321;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	483	GFSGPRCQ	490
DB	1237	GFSGPRCQ	1244

RESULT 15
 NTC2_MOUSE
 ID NTC2_MOUSE STANDARD; PRT; 2470 AA.
 AC O35516; Q06008; Q0941;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch
 DE B).
 GN NOTCH2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=CSHEL/6; TISSUE=Thymus;
 RA Hamada Y., Higuchi M., Tsujimoto Y.;
 RT "Complete amino acid sequence and multiform transcripts encoded by a
 RT single copy of mouse Notch2 gene";
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RC SEQUENCE OF 316-1518 FROM N.A.
 RP STRAIN=CSHEL/6 X CBA; TISSUE=Embryo;
 RX MEDLINE=93178563; PubMed=8440332;
 RA Lardelli M., Lendahl U.;
 RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
 RT wide variety of tissues";
 RL Exp. Cell Res. 204:364-372(1993).
 RN [3]
 RP SEQUENCE OF 1765-2153 FROM N.A.

RX MEDLINE=97075110; PubMed=8917536;
 RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,
 RA Martin D.I.;
 RT "Inhibition of granulocytic differentiation by mNotch1";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).
 RN [4]
 RP FUNCTION.
 RX MEDLINE=99396706; PubMed=10393120;
 RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,
 RA Tsujimoto Y.;
 RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
 RT embryonic lethality";
 RL Development 126:3415-3424(1999).
 RN [5]
 RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
 RX MEDLINE=95333893; PubMed=7609614;
 RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
 RT "Differential expression of Notch1 and Notch2 in developing and adult
 RT mouse brain";
 RL Brain Res. Mol. Brain Res. 29:263-272(1995).
 RN [6]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 RT proteolysis";
 RL J. Biol. Chem. 276:40269-40273(2001).
 RN [7]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). May play an essential role in
 CC postimplantation development, probably in some aspect of cell
 CC specification and/or differentiation.
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=O35516-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=O35516-2; Sequence=VSP_001405;
 CC Note=No experimental confirmation available;
 CC -!- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
 CC neuroepithelia, somites, optic vesicles and branchial arches, but
 CC not heart.
 CC -!- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,
 CC the postnatal ependymal cells, and the choroid plexus throughout
 CC embryonic and postnatal development.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC -!- PTM: Phosphorylated.

CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 35 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 6 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D32210; BAA22094.1; -
 CC EMBL; X68279; CAA48340.1; -
 CC EMBL; U31881; AAC52924.1; -
 CC FRR; A49175; A49175.
 CC HSP; P16109; IFSB.
 CC MGD; MGI:97364; Notch2.
 CC GO; GO:0005887; C:integral to plasma membrane; IC.
 CC GO; GO:0005515; F:protein binding; IPI.
 CC GO; GO:0002011; P:morphogenesis of an epithelial sheet; IMP.
 CC GO; GO:0007219; P:N signaling pathway; IC.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000152; Asx_hydroxyl_s.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR001438; EGF_II.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR008297; Notch.
 CC InterPro; IPR000800; Notch_dom.
 CC Pfam; PF00023; ANK; 6.
 CC Pfam; PF00008; EGF; 34.
 CC Pfam; PF00066; notch; 2.
 CC PIRSF; PIRSF002279; Notch; 1.
 CC PRINTS; PR00101; EGFLOOD.
 CC PRINTS; PR00011; EGFAMININ.
 CC PRINTS; PR01452; NOTCH.
 CC SMART; SM00248; ANK; 6.
 CC SMART; SM00179; EGF_Ca; 23.
 CC SMART; SM00004; NL; 3.
 CC PROSITE; PS0297; ANK REP REGION; 1.
 CC PROSITE; PS0088; ANK REPEAT; 4.
 CC PROSITE; PS0010; ASX_HYDROXYL; 22.
 CC PROSITE; PS00022; EGF_1; 33.
 CC PROSITE; PS01186; EGF_2; 27.
 CC PROSITE; PS00026; EGF_3; 35.
 CC PROSITE; PS01187; EGF_Ca; 22.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation;
 CC Alternative splicing.
 CC SIGNAL; 1 25
 CC CHAIN; 26 2470
 CC CHAIN; 1656 2470
 CC CHAIN; 1697 2470
 CC DOMAIN; 26 1677
 CC DOMAIN; 1678 1697
 CC DOMAIN; 1699 2470
 CC DOMAIN; 26 63
 CC DOMAIN; 64 102
 CC DOMAIN; 105 143
 CC DOMAIN; 144 180
 CC DOMAIN; 182 219
 CC DOMAIN; 221 256
 CC DOMAIN; 258 294
 CC DOMAIN; 296 334
 CC DOMAIN; 336 372
 CC DOMAIN; 373 411
 CC DOMAIN; 413 452
 CC DOMAIN; 454 490
 CC DOMAIN; 492 528
 CC POTENTIAL.
 CC NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
 CC NOTCH EXTRACELLULAR TRUNCATION.
 CC NOTCH INTRACELLULAR DOMAIN.
 CC EXTRACELLULAR (POTENTIAL).
 CC POTENTIAL.
 CC CYTOPLASMIC (POTENTIAL).
 CC EGF-LIKE 1.
 CC EGF-LIKE 2.
 CC EGF-LIKE 3.
 CC EGF-LIKE 4.
 CC EGF-LIKE 5 (INCOMPLETE).
 CC EGF-LIKE 6 (POTENTIAL).
 CC EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 CC EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 CC EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 CC EGF-LIKE 10.
 CC EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 CC EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 CC EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN	530	566	EGF-LIKE 14,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	568	603	EGF-LIKE 15,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	605	641	EGF-LIKE 16,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	643	678	EGF-LIKE 17,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	680	716	EGF-LIKE 18,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	718	753	EGF-LIKE 19,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	755	791	EGF-LIKE 20,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	793	829	EGF-LIKE 21,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	831	869	EGF-LIKE 22,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	871	907	EGF-LIKE 23,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	909	945	EGF-LIKE 24,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	947	983	EGF-LIKE 25,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	985	1021	EGF-LIKE 26,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	1023	1059	EGF-LIKE 27,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	1061	1097	EGF-LIKE 28,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	1099	1145	EGF-LIKE 29,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	1147	1183	EGF-LIKE 30,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	1185	1221	EGF-LIKE 31,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	1223	1260	EGF-LIKE 32,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	1262	1300	EGF-LIKE 33,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	1302	1345	EGF-LIKE 34,	CALCIUM-BINDING	(POTENTIAL).
FT DOMAIN	1372	1410	EGF-LIKE 35,	CALCIUM-BINDING	(POTENTIAL).
FT REPEAT	1418	1454	LIN/NOTCH 1.		
FT REPEAT	1501	1533	LIN/NOTCH 2.		
FT REPEAT	1825	1869	ANK 1.		

Query Match 1.1%; Score 8; DB 1; Length 2470;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 426 PCQHGATC 433

DB 1155 PCQHGATC 1162

Search completed: March 9, 2004, 17:28:34
 Job time: 19.0517 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:22:43 ; Search time 61.7389 Seconds
(without alignments)
3602.917 Million cell updates/sec

Title: US-10-006-011A-3

Perfect score: 705

Sequence: 1 EIKITFRPDSADGMLLYNQ.....QPLDLQHAQAGANTRPCPS 705

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	1.8	68	11	008591
2	10	1.4	53	13	Q9DFLO
3	9	1.3	1208	11	Q80YA8
4	9	1.3	2516	11	Q7TQ52
5	9	1.3	2526	11	Q7TQ51
6	9	1.3	2531	11	Q8K428
7	9	1.3	2531	11	Q7TQ50
8	9	1.3	3367	5	Q9XZC9
9	9	1.3	3375	5	Q8IP51
10	8	1.1	97	5	Q8I718
11	8	1.1	219	11	Q8BMI5
12	8	1.1	244	3	Q9P6X9
13	8	1.1	302	16	Q82W41
14	8	1.1	308	2	Q842M1
15	8	1.1	320	16	Q8G4F5
16	8	1.1	382	11	Q8K1E3

17	8	1.1	388	5	Q7YWC2
18	8	1.1	430	16	Q92VG0
19	8	1.1	487	5	Q8MSX5
20	8	1.1	487	16	Q9FBM8
21	8	1.1	489	4	Q8N1J0
22	8	1.1	494	16	Q81XS5
23	8	1.1	494	16	Q816E3
24	8	1.1	497	2	Q7X2C8
25	8	1.1	497	16	Q8YWD5
26	8	1.1	615	13	O57409
27	8	1.1	642	13	P79941
28	8	1.1	642	13	Q7ZKT4
29	8	1.1	659	11	Q7TQI9
30	8	1.1	662	4	Q9H8X6
31	8	1.1	664	13	Q9IAT6
32	8	1.1	702	4	Q9HA02
33	8	1.1	731	4	Q9HD85
34	8	1.1	731	4	Q96AQ6
35	8	1.1	748	10	Q9FRR2
36	8	1.1	816	11	Q70474
37	8	1.1	858	11	Q8R145
38	8	1.1	886	3	O14260
39	8	1.1	1016	16	Q8PKK4
40	8	1.1	1290	13	Q9W6E1
41	8	1.1	1551	5	Q9NGV4
42	8	1.1	1697	12	O8JUG7
43	8	1.1	1697	12	O8JUL0
44	8	1.1	2146	5	Q9VC97
45	8	1.1	2447	13	O13149

ALIGNMENTS

RESULT 1

008591 PRELIMINARY; PRT; 68 AA.
AC O08591;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Parlecan (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Pyke C., Kristensen P., Ostergaard P.B., Oturai P.S., Romer J.;
RT "Proteoglycan expression in the normal rat kidney."
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U75305; AAB51124.1;
DR InterPro; IPR001985; ConA like lec_gl.
DR InterPro; IPR001791; Laminin_G_1.
DR Pfam; PF00054; laminin_G_1.
DR PROSITE; PS00025; LAM_G_DOMAIN; 1.
FT NON_TER 1
FT NON_TER 68
SQ SEQUENCE 68 AA; 7395 MW; 5868E45D8A7083E0 CRC64;

Query Match 1.8%; Score 13; DB 11; Length 68;
Best Local Similarity 100.0%; Pred. No. 4.5e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 533 FRSRLPEVPETIE 545

Db 21 FRSRLPEVPETIE 33

RESULT 2

Q9DFLO PRELIMINARY; PRT; 53 AA.
ID Q9DFLO

```

AC Q9DFLO;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE S-adenosylmethionine synthase-like protein (Fragment).
OS Gillichthys mirabilis (long-jawed mudsucker).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Gobioidae;
OC Gobiidae; Gillichthys.
OX NCBI_TaxID=8222;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA MEDLINE=21117151; PubMed=11172064;
RA Gracey A.Y., Troll J.V., Somero G.N.;
RT "Hypoxia-induced gene expression profiling in the euryoxic fish
RL Gillichthys mirabilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:1993-1998(2001).
DR EMBL; AF286225; AAG13345.1; -.
FT NON_TER 1
SQ SEQUENCE 53 AA; 5541 MW; A7D6F0FP411CF46E CRC64;

Query Match 1.4%; Score 10; DB 13; Length 53;
Best Local Similarity 100.0%; Pred. No. 0.053;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 474 QGTRCLCLPG 483
DB 35 QGTRCLCLPG 44

RESULT 3
Q80YA8 PRELIMINARY; PRT; 1208 AA.
ID Q80YA8
AC Q80YA8
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE 593040A21 protein (Fragment).
GN 593040A21.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.X.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.B., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Ramey J., Heiton E., Kettner M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RE and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;

```

```

RL Submitted (JAN-2003) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC043114; AAH43114.1; -.
DR GO; GO:000509; F:calcium ion binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001841; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001791; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00008; EGF; 13.
DR PRINTS; PR00010; EGFLOOD.
DR PRINTS; PR00011; EGFLAMININ.
DR SMART; SM00181; EGF; 15.
DR SMART; SM00179; EGF_CA; 13.
DR SMART; SM00282; LamG; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 8.
DR PROSITE; PS00022; EGF_1; 14.
DR PROSITE; PS01186; EGF_2; 9.
DR PROSITE; PS01187; EGF_CA; 5.
DR PROSITE; PS50025; LAM_G_DOMAIN; 2.
FT NON_TER 1
SQ SEQUENCE 1208 AA; 126978 MW; FDF2DF2F0B2F198 CRC64;

Query Match 1.3%; Score 9; DB 11; Length 1208;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 EPCLHGTC 473
DB 295 EPCLHGTC 303

RESULT 4
Q7TQ52 PRELIMINARY; PRT; 2516 AA.
ID Q7TQ52
AC Q7TQ52
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Transmembrane receptor Notch1 B.
GN NOTCH1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C.B-17; TISSUE=Thymus;
RA Tsuji H., Ishii-Onba H., Ukai H., Katsube T., Ogiu T.;
RT "Radiation-induced deletions in the 5' end region of Notch1 lead to
RT the formation of truncated proteins and are involved in the
RT development of mouse thymic lymphomas.";
RL Carcinogenesis 24:1-12(2003).
DR EMBL; AB100603; BAC77038.1; -.
KW Receptor; Transmembrane.
SQ SEQUENCE 2516 AA; 269177 MW; 17FD72740EBD6E35 CRC64;

Query Match 1.3%; Score 9; DB 11; Length 2516;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGTCQ 474
DB 1014 PCLHGTCQ 1022

RESULT 5
Q7TQ51 PRELIMINARY; PRT; 2526 AA.
ID Q7TQ51
AC Q7TQ51;

```

```

DT 01-OCT-2003 (TREMELrel. 25, Created)
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Transmembrane receptor Notch1 D.
GN NOTCH1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C.B-17; TISSUE=Thymus;
RA Tsuji H., Ishii-Ohba H., Ukai H., Katsube T., Ogii T.;
RT "Radiation-induced deletions in the 5' end region of Notch1 lead to
RT the formation of truncated proteins and are involved in the
RT development of mouse thymic lymphomas.";
RL Carcinogenesis 24:1-12(2003).
DR EMBL; AB100603; BAC77040.1; -.
KW Receptor; Transmembrane.
SQ SEQUENCE 2526 AA; 270583 MW; 017563FCE9703264 CRC64;

Query Match 1.3%; Score 9; DB 11; Length 2526;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGGTCQ 474
DB 1024 PCLHGGTCQ 1032

RESULT 6
Q8K428 PRELIMINARY; PRT; 2531 AA.
AC Q8K428;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Notch 1 protein.
GN NOTCH1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Thymus;
RX MEDLINE=95044925; PubMed=7956822;
RA Nye J.S., Kopan R., Axel R.;
RT "An activated Notch suppresses neurogenesis and myogenesis but not
RT gliogenesis in mammalian cells.";
RL Development 120:2421-2430(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Thymus;
RX MEDLINE=22119593; PubMed=12123574;
RA Folitz D.R., Santiago M.C., Berechid B.E., Nye J.S.;
RT "Glycogen Synthase Kinase-3beta Modulates Notch Signaling and
RT Stability.";
RL Curr. Biol. 12:1006-1011(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Thymus;
RA Schellin K.A., Paulley A.M., Nye J.S.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF508809; AAM28905.1; -.
DR MGD; MGI.97363; Notch1.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0030154; P:cell differentiation; IMP.
DR GO; GO:0007386; P:compartment specification; IMP.
DR GO; GO:0045944; P:positive regulation of transcription from P. .; IDA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Aex_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.

```

```

DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 36.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00010; EGFBLLOOD.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00181; EGF; 37.
DR SMART; SM00179; EGF_CA; 35.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 35.
DR PROSITE; PS01186; EGF_2; 27.
DR PROSITE; PS01187; EGF_CA; 21.
DR PIRSF; PIRSF002279; Notch; 1.
KW ANK repeat; EGF-like domain; Repeat.
SQ SEQUENCE 2531 AA; 270819 MW; 7DB7E0DEF799D999 CRC64;

Query Match 1.3%; Score 9; DB 11; Length 2531;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGGTCQ 474
DB 1029 PCLHGGTCQ 1037

RESULT 7
Q7TQ50 PRELIMINARY; PRT; 2531 AA.
AC Q7TQ50;
DT 01-OCT-2003 (TREMELrel. 25, Created)
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Transmembrane receptor Notch1.
GN NOTCH1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C.B-17; TISSUE=Thymus;
RA Tsuji H., Ishii-Ohba H., Ukai H., Katsube T., Ogii T.;
RT "Radiation-induced deletions in the 5' end region of Notch1 lead to
RT the formation of truncated proteins and are involved in the
RT development of mouse thymic lymphomas.";
RL Carcinogenesis 24:1-12(2003).
DR EMBL; AB100603; BAC77040.1; -.
KW Receptor; Transmembrane.
SQ SEQUENCE 2531 AA; 270832 MW; 97C91F69BABF02BF CRC64;

Query Match 1.3%; Score 9; DB 11; Length 2531;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGGTCQ 474
DB 1029 PCLHGGTCQ 1037

RESULT 8
Q9AZC9 PRELIMINARY; PRT; 3367 AA.
ID Q9AZC9

```

AC QXZC9;
 DT 01-NOV-1999 (T-EMBLrel. 12, Created)
 DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Laminin ALPHA1.2 (SYMBOL=WB).
 GN WB OR WING BLISTER OR CGI5288.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=99207061; PubMed=10189378;
 RA Martin D., Zusman S., Li X., Williams E.L., Khare N., Darocha S.,
 RA Chiquet-Ehrismann R., Baumgartner S.,
 RT "wing blister, a new Drosophila laminin alpha chain required for cell
 RT adhesion and migration during embryonic and imaginal development.",
 RL J. Cell Biol. 145:191-201(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=9403001; PubMed=10471707;
 RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
 RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
 RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
 RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
 RA Celnik S., Rubin G.M.,
 RT "An exploration of the sequence of a 2.9-Mb region of the genome of
 RT Drosophila melanogaster: the Adh region.",
 RL Genetics 153:179-219(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RA Celnik S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
 RA Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
 RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
 RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
 RA Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
 RA Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
 RA Sethi H., Snir E., Swirskas R.R., Wan K.H., Weinburg T., Zhang R.,
 RA Zieran L.L., Rubin G.M.,
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF135118; AAD31714.1;
 DR EMBL; AE003409; AAF44861.1;
 DR HSP; P00740; 1EDM.
 DR FlyBase; FBgn004002; wb.
 DR GO; GO:0007267; F:cell-cell signaling; NAS.
 DR InterPro; IPR001589; Actbind_actinin.
 DR InterPro; IPR008985; ConA_like_lec_gl.
 DR InterPro; IPR006209; EGF like.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR008211; LamNT.
 DR Pfam; PF00052; laminin_B; 2.
 DR Pfam; PF00053; laminin_EGF; 14.
 DR Pfam; PF00054; laminin_G; 4.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR SMART; SM00180; EGF_Lam; 16.
 DR SMART; SM00282; LamG; 5.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00019; ACTININ_1; 1.
 DR PROSITE; PS00022; EGF_1; 14.
 DR PROSITE; PS01186; EGF_2; 4.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 16.
 DR PROSITE; PS00025; LAM_G_DOWNIN; 4.
 KW Laminin EGF-like domain.
 SQ SEQUENCE 3367 AA; 374097 MW; EB125654B1BC1511 CRC64;

Query Match

1.3%; Score 9; DB 5; Length 3367;

Best Local Similarity 100.0%; Pred. No. 26;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 603 NDGEHRTV 611
 |||||
 DB 3081 NDGEHRTV 3089

RESULT 9

Q8IP51 PRELIMINARY; PRT; 3375 AA.
 ID Q8IP51
 AC Q8IP51;
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE CGI5288-PB.
 GN WB OR CGI5288.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abail J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fodor C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Follard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhou X., Zhou S., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celnik S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banjon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., K.
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.
RA Chan J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Eisen M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Perteira M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum".
RL Nature 419:498-511(2002).
RL EMBL; AEO14824; AAN37127.1; -.
DR GO; GO:0000214; C:trna-intron endonuclease complex; IEA.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0000213; F:trna-intron endonuclease activity; IEA.
DR GO; GO:0006388; P:trna splicing; IEA.
DR InterPro; IPR006577; trna_int_endo_C.
DR Pfam; PF01974; trna_int_endo; 1.
DR Endonuclease.
KW
SQ SEQUENCE 97 AA; 11640 MW; D27B07F8A1360441 CRC64;

Query Match 1.1%; Score 8; DB 5; Length 97;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 143 EIVFHDLN 150
| | | | | | | | | |
DB 9 EIVFHDLN 16

RESULT 11
Q8BM15
ID Q8BM15 PRELIMINARY; PRT; 219 AA.
AC Q8BM15;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Weakly similar to neurogenic locus notch 3 protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Forelimb;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs".
RL Nature 420:563-573 (2002).
RL EMBL; AK031068; BAC27237.1; -.
DR PIR; P70633; P70633.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR000152; Asx hydroxyl_S.
DR InterPro; IPR000742; EGF 2.
DR InterPro; IPR001881; EGF-Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.
DR Pfam; PF00008; EGF; 2.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00179; EGF CA; 2.
DR PROSITE; PS00010; ASX HYDROXYL; 1.
DR PROSITE; PS00022; EGF 1; 4.
DR PROSITE; PS01186; EGF 2; 2.
FT NON TER 1

SQ SEQUENCE 219 AA; 23098 MW; 9F2C4BE237F9B522 CRC64;

Query Match 1.1%; Score 8; DB 11; Length 219;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 482 PGFSGPRC 489
DB 56 PGFSGPRC 63

RESULT 12

Q9F6X9 PRELIMINARY; PRT; 244 AA.

ID Q9P6X9;
AC Q9P6X9;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Related to putative lipid binding protein TFS1.
GN 13E11.370.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL353820; CAB98605.2; -;
DR HSP; F30086; 1BD9.
DR InterPro; IPR008914; PEPF.
DR Pfam; PF01161; PBP; 1.
SQ SEQUENCE 244 AA; 26745 MW; EEB0DC9E6245012A CRC64;

Query Match 1.1%; Score 8; DB 3; Length 244;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 LAPVNGTS 97
DB 186 LAPVNGTS 193

RESULT 13

Q82W41 PRELIMINARY; PRT; 302 AA.

ID Q82W41;
AC Q82W41;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Phosphadenosine phosphosulfate reductase (EC 2.7.7.4).
GN CYSD OR NE0856.
OS Nitrosomonas europaea.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_TaxID=915;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19718 / IFO 14298;
RX MEDLINE=22586410; PubMed=12700255;
RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Savavedra-Soto L.A.,
RA Arclero D.M., Hommes N.G., Whittaker M.M., Atp D.J.;
RT "Complete genome sequence of the ammonia-oxidizing bacterium and
obligate chemolithoautotroph Nitrosomonas europaea.";
RL J. Bacteriol. 185:2759-2773 (2003).
DR EMBL; BX321859; CAD84767.1; -;
GO; GO:0016779; F:nucleotidyltransferase activity; IEA.

DR GO; GO:0004781; F:sulfate adenylyltransferase (ATP) activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002500; PAPS_reduct.
DR Pfam; PF01507; PAPS_reduct; 1.
KW Nucleotidyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 302 AA; 34973 MW; 5499EAC71EF9F7DD CRC64;

Query Match 1.1%; Score 8; DB 16; Length 302;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 VLLFSGGK 286
DB 32 VLLFSGGK 39

RESULT 14

Q842M1 PRELIMINARY; PRT; 308 AA.

ID Q842M1;
AC Q842M1;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE CYSD (EC 2.7.7.4).
GN CYSD.
OS Rhodococcus sp. DS7.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=227347;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DS7;
RA Vitale E., Brunelli E., de Ferra F.;
RT "Effects of cybNC mutations on the desulphurization of organosulphur
compounds in Rhodococcus sp. strain DS7.";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY244763; AAO89190.1; -;
DR GO; GO:0016779; F:nucleotidyltransferase activity; IEA.
DR GO; GO:0004781; F:sulfate adenylyltransferase (ATP) activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002500; PAPS_reduct.
DR Pfam; PF01507; PAPS_reduct; 1.
KW Nucleotidyltransferase; Transferase.
SQ SEQUENCE 308 AA; 35136 MW; AE84C09A9D453F7C CRC64;

Query Match 1.1%; Score 8; DB 2; Length 308;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 VLLFSGGK 286
DB 37 VLLFSGGK 44

RESULT 15

Q8G4F5 PRELIMINARY; PRT; 320 AA.

ID Q8G4F5;
AC Q8G4F5;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Probable ribokinase.
GN BL1431.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787;

RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
 RA Pessi G., Zwaalen M.-C., Desiere F., Bork P., Delley M.,
 RA Pridmore R.D., Arigoni F.,
 RT "The genome sequence of Bifidobacterium longum reflects its adaptation
 to the human gastrointestinal tract.",
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
 DR EMBL; AF014773; AAN25228.1; -.
 DR GO; GO:0004747; F:ribokinase activity; IEA.
 DR GO; GO:0006014; P:D-ribose metabolism; IEA.
 DR InterPro; IPR002173; PfKB.
 DR InterPro; IPR002139; Ribokinase.
 DR Pfam; PF00294; pfkB; 1.
 DR PRINTS; PRO0920; RIBOKINASE.
 DR PROSITE; PS00583; PFKB_KINASES_1; 1.
 KW Kinase; Complete proteome.
 SQ SEQUENCE 320 AA; 33130 MW; 28F44619036EFC6B CRC64; .

Query Match 1.1%; Score 8; DB 16; Length 320;
 Best Local Similarity 100.0%; Pred.No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 GWHVSA 332

DB 249 GWHVSA 256

Search completed: March 9, 2004, 17:30:33
 Job time : 64.739 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:15:33 ; Search time 24.1076 Seconds
(without alignments)
2813.016 Million cell updates/sec

Title: US-10-006-011A-3

Perfect score: 3825

Sequence: 1 EIKITFRPDSADGMLLYNQ.....QLDLQHQAGANTPCPS 705

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:*

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3825	100.0	4391	2 A38096	perlecan precursor
2	3389.5	88.6	3707	2 S18252	heparan sulfate pr
3	859	22.5	1959	1 AGRT	agrin - rat
4	796.5	20.8	1955	1 AGCH	agrin precursor -
5	667	17.4	1328	2 T43060	agrin - electric r
6	511	13.4	3375	2 T19821	hypothetical prote
7	421	11.0	1715	2 C40228	neurexin II-alpha
8	419.5	11.0	1531	2 T42218	slit-1 protein hom
9	405.5	10.6	1530	2 T45344	neurexin I-alpha
10	403.5	10.5	1507	2 A40228	neurexin I-alpha p
11	387.5	10.1	1025	2 T42626	secreted leucine-r
12	387	10.1	1523	2 T13953	MEGF5 protein - ra
13	383	10.0	5147	1 IJFFTM	cadherin-related t
14	382	10.0	1438	2 A48216	neurexin III-alpha
15	382	10.0	1471	2 B48218	neurexin III-alpha
16	382	10.0	1578	2 T48216	neurexin III-alpha
17	362.5	9.5	4307	2 T20721	hypothetical prote
18	357	9.3	1751	1 MMHUNH	laminin alpha-2 ch
19	353.5	9.2	3084	1 MMW5A	laminin alpha-1 ch
20	353	9.2	3075	2 S14458	laminin alpha-1 ch
21	349	9.1	3097	2 T00021	DN-cadherin - frui
22	347	9.1	3106	1 S23868	laminin alpha-2 ch
23	344.5	9.0	601	2 T22025	hypothetical prote
24	344.5	9.0	601	2 D89711	protein F40E10.4 [
25	340	8.9	2531	2 T31070	slit protein - se
26	338.5	8.8	1469	2 B36665	slit protein 2 pre
27	333	8.7	1480	2 A36665	slit protein 1 pre
28	326	8.5	3672	2 T23433	hypothetical prote
29	319	8.3	2471	2 A49128	cell-fate determin

30 317 8.3 3704 2 T37316
31 316 8.3 2321 2 S78549
32 314.5 8.2 1203 2 A49175
33 311 8.1 2318 2 S45306
34 311 8.1 3034 2 T14119
35 304 7.9 2809 2 T30213
36 302.5 7.9 1220 2 A56136
37 302.5 7.9 2555 2 A40043
38 299.5 7.8 2524 2 A35844
39 298 7.8 4351 2 T00252
40 296 7.7 2531 2 S18188
41 296 7.7 2531 2 A46019
42 292 7.6 861 2 A48825
43 291 7.6 1964 2 T03059
44 287.5 7.5 2352 2 T30201
45 284.5 7.4 2703 1 A24420

ALIGNMENTS

RESULT 1

A38096
perlecan precursor - human
N; Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate proteoglycan
C; Species: Homo sapiens (man)
C; Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #text change 05-Nov-1999
C; Accession: A38096; S19256; S77946; A41059; A40306; B33625; A36225; A41736
R; Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.
J. Biol. Chem. 267, 8544-8557, 1992
A; Title: Primary structure of the human heparan sulfate proteoglycan from basement membr.
tor, laminin, neural cell adhesion molecules, and epidermal growth factor.
A; Reference number: A38096; MUID:92235084; PMID:1569102
A; Accession: A38096
A; Molecule type: mRNA
A; Residues: 1-4391 <MUR>
A; Cross-references: GB:M85289; NID:gl84426; PIDN:AAA52700.1; PID:gl84427
R; Kallunki, P.; Tryggvason, K.
J. Cell Biol. 116, 559-571, 1992
A; Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD pro:
eii adhesion molecules, and epidermal growth factor.
A; Reference number: A41736; MUID:92112994; PMID:1730768
A; Accession: S19256
A; Molecule type: mRNA
A; Residues: 1-57, 'D', 'S', '9', '434', 'A', '436', 'E', '438-449', 'Q', '451-502', 'A', '503-792', 'K', '794-908', 'R',
71-2979, 'H', '2981-2994', 'G', '2996-3167', 'T', '3169-3240', 'R', '3242-3436', 'R', '3428-3631', 'Q', '3633-3.
A; Cross-references: EMBL:X62515
R; Tryggvason, K.
submitted to the EMBL Data Library, October 1991
A; Reference number: S77946
A; Accession: S77946
A; Molecule type: mRNA
A; Residues: 1-57, 'D', 'S', '9', '434', 'A', '436', 'E', '438-449', 'Q', '451-502', 'A', '503-792', 'K', '794-908', 'R',
71-2979, 'H', '2981-2994', 'G', '2996-3167', 'T', '3169-3240', 'R', '3242-3436', 'R', '3428-3631', 'Q', '3633-4.
A; Cross-references: EMBL:X62515; NID:q29469; PIDN:CAA44373.1; PID:q29470
R; Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestilae, M.; Shows, T.B.; Tryggvason, K.
Genomics 11, 389-396, 1991
A; Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the g
A; Reference number: A41059; MUID:92120660; PMID:1685141
A; Accession: A41059
A; Molecule type: mRNA
A; Residues: 'R', '892-908', 'R', '910-1101', 'L', '1103-1132', 'L', '1134-1221', 'L', '1223-1397' <KA2>
A; Cross-references: GB:S76436; NID:q243370; PIDN:AA821121.1; PID:q243371
R; Dodge, G.R.; Kovalsky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo, J
Genomics 10, 673-680, 1991
A; Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellula:
A; Reference number: A40306; MUID:91365376; PMID:1679749
A; Accession: A40306
A; Molecule type: mRNA
A; Residues: 1018-1405, 'G', '1407-1409', 'G', '1411-1465' <DOD>
A; Cross-references: GB:M64283; NID:gl84424; PIDN:AAA52699.1; PID:gl84425
R; Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Cassiman, J.J.; van den
J. Cell Biol. 109, 3199-3211, 1989

A;Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclonal antibodies.
 A;Reference number: A33625; MUID:90078352; PMID:2687294
 A;Accession: B33625
 A;Molecule type: protein
 A;Residues: 1379-1384, 'X', 1386-1388, 'X', 1390-1398 <HE2>
 A;Accession: A33625
 A;Molecule type: protein
 A;Residues: 2166-2171, 'X', 2173-2175, 'X', 2177-2185 <HE3>
 A;Note: peptide potentially matches four different regions of sequence shown
 C;Genetic: GDB:HSFG2
 A;Gene: GDB:HSFG2
 A;Cross-references: GDB:126372; OMIM:142461
 A;Map position: ip36.1-1p36.1
 C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repeats
 C;Keywords: Chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmembrane
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-4391/Product: perlecan #status predicted <MAT>
 F:22-193/Domain: I <DOM1>
 F:194-530/Domain: II <DOM2>
 F:199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F:285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F:368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
 F:531-1676/Domain: III <DOM3>
 F:1159-1206/Domain: laminin-type EGF-like homology <LEG>
 F:1563-1610/Domain: laminin-type EGF-like homology <LEG>
 F:1613-1668/Domain: laminin-type EGF-like homology <LEG>
 F:1677-3686/Domain: IV <DOM4>
 F:2007-2034/Domain: transmembrane #status predicted <TRM>
 F:3687-4391/Domain: V <DOM5>
 F:3845-3880/Domain: EGF homology <EGF1>
 F:3888-3921/Domain: EGF homology <EGF>
 F:3953-4106/Domain: laminin G repeat homology <EGF2>
 F:4149-4151/Region: motor neuron attachment (L-R-E) motif
 F:4299-4301/Region: motor neuron attachment (L-R-E) motif
 F:657,71,76/Binding site: heparan sulfate (Ser) (covalent) #status predicted
 F:89,554,1755,2121,3072,3105,3279,3780,3836,4068/Binding site: carbohydrate (Asn) (covalent)
 F:2995,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 3825; DB 2; Length 4391;
 Best Local Similarity 100.0%; Pred. No. 5.4e-237;
 Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKIIFRDSADGMLLYNGKRVGSPNTLANRQPDFISGLVGRPEFRFDAGSGMATI 60
 DB 3687 EKIIFRDSADGMLLYNGKRVGSPNTLANRQPDFISGLVGRPEFRFDAGSGMATI 3746
 QY 61 RHPTPLALGHFHTVTLRLSLTQGSLLVGLDAPVNGTSQKFGQLDNEELVLCGYPDYGA 120
 DB 3747 RHPTPLALGHFHTVTLRLSLTQGSLLVGLDAPVNGTSQKFGQLDNEELVLCGYPDYGA 3806
 QY 121 IPKAGLSGFTGCVRELAIQEEIVFHDNLNTAGISHCPTCRDRPQNGQCHDSSES 180
 DB 3807 IPKAGLSGFTGCVRELAIQEEIVFHDNLNTAGISHCPTCRDRPQNGQCHDSSES 3866
 QY 181 YVCVCPAGFTGRCRCHSHQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSLRCSEGV 240
 DB 3867 YVCVCPAGFTGRCRCHSHQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSLRCSEGV 3926
 QY 241 VTTSLSGAGSVLALPALTNTHTHLLDVEKPLAPDGVLLFSGKSGPVEDFVSLAMVG 300
 DB 3927 VTTSLSGAGSVLALPALTNTHTHLLDVEKPLAPDGVLLFSGKSGPVEDFVSLAMVG 3986
 QY 301 GHLEFRYELGSLAVLRSAEPLALGRHVRSAERLNKDGSLRVNNGRPVLRSPGKSGQL 360
 DB 3987 GHLEFRYELGSLAVLRSAEPLALGRHVRSAERLNKDGSLRVNNGRPVLRSPGKSGQL 4046
 QY 361 NLHTLLYLGVEPVSPLSPATNMAHFRGCVGVSVNGKRLDLTYSPFLSGQIGQCYDSS 420
 DB 4047 NLHTLLYLGVEPVSPLSPATNMAHFRGCVGVSVNGKRLDLTYSPFLSGQIGQCYDSS 4106

QY 421 PCRQPCQHGATCMPAGEYFQCLCRDGFKGDI,CEHEENPCQREPCCLHGCTCGTRCLC 480
 DB 4107 PCRQPCQHGATCMPAGEYFQCLCRDGFKGDI,CEHEENPCQREPCCLHGCTCGTRCLC 4166
 QY 481 LPGFSGPRCCQSGHGIAESDWHLESGGNDAPQGVAYFHDDGFLAFPHGVFSRLPEV 540
 DB 4167 LPGFSGPRCCQSGHGIAESDWHLESGGNDAPQGVAYFHDDGFLAFPHGVFSRLPEV 4226
 QY 541 PETIELEVRTSTASGLLLWQGVGVGAGQKQFISGLQDGHVLFVRYQLGSGEARLVSE 600
 DB 4227 PETIELEVRTSTASGLLLWQGVGVGAGQKQFISGLQDGHVLFVRYQLGSGEARLVSE 4286
 QY 601 PINDGEHVRVTALREGRGSIQVDGELVSGRSPGNVAVNAKGSVYIGAPDVALTGG 660
 DB 4287 PINDGEHVRVTALREGRGSIQVDGELVSGRSPGNVAVNAKGSVYIGAPDVALTGG 4346
 QY 661 RFSSGTTGCVKLVLSARPGAPPPQPLDLQHRAGANTRPCPS 705
 DB 4347 RFSSGTTGCVKLVLSARPGAPPPQPLDLQHRAGANTRPCPS 4391

RESULT 2
 S18252
 heparan sulfate proteoglycan - mouse
 N;Alternate names: perlecan
 C;Species: Mus musculus (house mouse)
 C;Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text_change 05-Nov-1999
 C;Accession: S18252; A31917; E31917; S66460
 R;Noonan, D.M.; Fuller, A.; Valente, P.; Cai, S.; Horigan, E.; Sasaki, M.; Yamada, Y.; Ha
 J. Biol. Chem. 266, 22939-22947, 1991
 A;Title: The complete sequence of perlecan, a basement membrane heparan sulfate proteogl
 adhesion molecule.
 A;Reference number: S18252; MUID:92078153; PMID:1744087
 A;Accession: S18252
 A;Molecule type: mRNA
 A;Residues: 1-3707 <NOO>
 A;Cross-references: EMBL:M77174; NID:G200295; PIDN:AAA39911.1; PID:G200296
 R;Noonan, D.M.; Horigan, E.A.; Ledbetter, S.R.; Vogeli, G.; Sasaki, M.; Yamada, Y.; Hass
 J. Biol. Chem. 263, 16379-16387, 1988
 A;Title: Identification of cDNA clones encoding different domains of the basement membra
 A;Reference number: A92680; MUID:89034110; PMID:2972708
 A;Accession: A31917
 A;Molecule type: mRNA
 A;Residues: 940-1601 <NO2>
 A;Cross-references: GB:J04054; NID:G200252; PIDN:AAA399899.1; PID:G200253
 A;Accession: B31917
 A;Molecule type: mRNA
 A;Residues: 1870-2600 <NO3>
 A;Cross-references: GB:J04055; NID:G200300; PIDN:AAA39912.1; PID:G200301
 R;Schulze, B.; Mann, K.; Battistutta, R.; Wiedemann, H.; Timpl, R.
 Eur. J. Biochem. 231, 551-556, 1995
 A;Title: Structural properties of recombinant domain III-3 of perlecan containing a glob
 A;Reference number: S66460; MUID:95377282; PMID:7649154
 A;Accession: S66460
 A;Molecule type: protein
 A;Residues: 1272-1274, 'X', 1276, 'X', 1278-1279 <SCH>
 C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
 C;Keywords: glycoprotein
 F:199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F:285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F:368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
 F:764-811/Domain: laminin-type EGF-like homology <LEG>
 F:1159-1206/Domain: laminin-type EGF-like homology <LEG>
 F:1563-1610/Domain: laminin-type EGF-like homology <LEG>
 F:1613-1668/Domain: laminin-type EGF-like homology <LEG>
 F:1677-3686/Domain: laminin-type EGF-like homology <LEG>
 F:3270-3423/Domain: EGF homology <EGF>
 F:3464-3492/Domain: laminin G repeat homology <LG2>
 F:1256,1891,2336,2394,2427/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 88.6%; Score 3389.5; DB 2; Length 3707;
 Best Local Similarity 88.9%; Pred. No. 3.9e-209;

Db 1686 VSLRGHQL-----TQEHVLRADVSPFADHPCQA----- 1716
 QY 454 CEHEENPCOLRPCLHGTC---OQT-RCILCLPGFSPRCQCGSGHGIAESDWHLESGG 509
 Db 1717 -----LGNPCNLGSCVPREATYECCLPGGSGHLCKEGL-----VKSVG 1757
 QY 510 NDAPOQYGAIFDDGFLAPGHVFSRSPV-----PEYI----- 544
 Db 1758 -----DLETAPDGRTYIEYLVNAVIESELNTEIPAPETLDSRALFSEKALQSN 1805
 QY 545 --ELEVRTSTAGLLMQGVEGAGQKDFISLGLQGHVFRYOLGSGEARLVSEDPPI 602
 Db 1806 HPELSIRTEATQGLVW-----LKAARADYNALAVDGHQLSLVDLGSQPVLRSTVKV 1861
 QY 603 NDEGHRVTVLREGRGSGIQVDEBELVNSGRSPGNVAVNAKGSVYIGAP--DVATLTGG 660
 Db 1862 NTRNLRLIRAHREHREGSLQVNEAPVTGSSPLGATQDLDGALWGLQKLPVQALPK 1921
 QY 661 RPSGGITGCVKNLVHSARPAGPPQPLDLOHRAQAGANTPCPS 705
 Db 1922 AVGTGFGVCLRDVVVGH-----ROLHLEDAVTKPELRPCPT 1958

RESULT 4
 AGCH

C:Species: Gallus gallus (chicken)
 C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
 R:Accession: JH0591; A38857; B38857; I50692
 R:Tsim, K.W.K.; Ruegg, M.A.; Escher, G.; Kroeger, S.; McMahan, U.J.
 Neuron 8, 677-689, 1992
 A:Title: cDNA that encodes active agrin.
 A:Reference number: JH0591; MUID:92232297; PMID:1314620
 A:Accession: JH0591
 A:Molecule type: mRNA
 A:Residues: 1-1955 <TS1>
 A:Cross-references: GB:M94271; NID:G2111120; PIDN:AAA48585.1; PID:G2111121
 A:Experimental source: brain
 R:Ruegg, M.A.; Tsim, K.W.K.; Horton, S.E.; Kroeger, S.; Escher, G.; McMahan, U.J.
 Neuron 8, 691-699, 1992
 A:Title: The agrin gene codes for a family of basal lamina proteins that differ in function
 A:Reference number: A38857; MUID:92232298; PMID:1314621
 A:Contents: alternative splicing
 A:Accession: A38857
 A:Molecule type: mRNA
 A:Residues: 1132-1783; 1795-1955 <RU2>
 A:Cross-references: GB:M97371
 A:Accession: B38857
 A:Molecule type: mRNA
 A:Residues: 1221-1647; 1652-1783; 1794-1955 <RU3>
 A:Cross-references: GB:M97372
 A:Note: translation of the nucleotide sequence is not complete
 R:Thomas, W.S.; O'Dowd, D.K.; Smith, M.A.
 Dev. Biol. 158, 523-535, 1993
 A:Title: Developmental expression and alternative splicing of chick agrin RNA.
 A:Reference number: I50692; MUID:93345745; PMID:8393816
 A:Accession: I50692
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 'SHLSNEIPA', 1784-1795 <THO>
 A:Cross-references: EMBL:U07271; NID:G459665; PIDN:AAA16788.1; PID:G459666
 C:Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine receptors
 C:Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repeat
 C:Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
 F:1-38/Domain: signal sequence #status predicted <MAT>
 F:39-1955/Product: agrin #status predicted <AG1>
 F:39-1783; 1795-1955/Product: agrin-related protein 1 #status predicted <AG2>
 F:39-1647; 1652-1783; 1794-1955/Product: agrin-related protein 2 #status predicted <AG3>
 F:77-126/Domain: Kazal proteinase inhibitor homology <KPI1>
 F:152-201/Domain: Kazal proteinase inhibitor homology <KPI2>
 F:225-273/Domain: Kazal proteinase inhibitor homology <KPI3>
 F:295-344/Domain: Kazal proteinase inhibitor homology <KPI4>

F:370-418/Domain: Kazal proteinase inhibitor homology <KPI5>
 F:435-483/Domain: Kazal proteinase inhibitor homology <KPI6>
 F:500-548/Domain: Kazal proteinase inhibitor homology <KPI7>
 F:584-633/Domain: Kazal proteinase inhibitor homology <KPI8>
 F:675-726/Domain: laminin-type EGF-like homology <LE1>
 F:729-773/Domain: laminin-type EGF-like homology <LE2>
 F:801-851/Domain: Kazal proteinase inhibitor homology <KPI9>
 F:856-995/Region: serine/threonine-rich
 F:1150-1219/Region: serine/threonine-rich
 F:1233-1264/Domain: EGF homology <EG1>
 F:1294-1448/Domain: laminin G repeat homology <LGI>
 F:1429-1431/Region: motor neuron attachment (L-R-E) motif
 F:1450-1482/Domain: EGF homology <EG2>
 F:1489-1521/Domain: EGF homology <EG3>
 F:1560-1711/Domain: laminin G repeat homology <LG2>
 F:1718-1751/Domain: laminin G repeat homology <LG3>
 F:1803-1955/Domain: laminin G repeat homology <LG4>
 F:185-105; 94-126; 160-180; 189-201; 233-252; 241-273; 304-323; 312-344; 378-397; 386-418; 443-462;
 1482; 1489-1500; 1494-1510; 1512-1521/Disulfide bonds: #status predicted
 F:390; 659; 764; 814/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 20.8%; Score 796.5; DB 1; Length 1955;
 Best Local Similarity 30.6%; Pred. No. 4; 7e-43;
 Matches 221; Conservative 99; Mismatches 271; Indels 131; Gaps 25;

QY 2 IKITFRPDSAGDMLLYNGKRVPGSPNTLANRQDFISFLVGGRPFFPDAGSGMATIR 61
 Db 1294 IAMEFRATELSGLLYNGO-----NRKDFISLALVGGFVELRFTNGSGTGVIT 1342
 QY 62 HPTFLALGHFHTVTLRLSLTQSLVLDLAPVNGTSQKFGQGLDNEELYLGYP-DYGA 120
 Db 1343 SKRVPEPKWHQLVNNRRSGMLAV-DEHVSGESPTGTDGLNLDLTFVGAGEDQWA 1401
 QY 121 I--PKAGLSGFGICVR-----ELRQGEIVFHDLNLTAHGISHC--PTCRDRPC 167
 Db 1402 VVAERTAAATVGLKGSIRLLDVNNQVYDLREKGSVLY-----GSGVGCGNDPCHPNC 1455
 QY 168 QNGGQCHDESSVYVCVPAGTGRSCHESSQALHCHPEACGPDATCVNRPDGRGYTCRH 227
 Db 1456 HIGASCHVKEAEHFCULHSYTGTC--DERNCPDPTCHISATCLVLPFG-GAMCACP 1513
 QY 228 LGRSLRCEEGTVT-----TPSLGAGSYLALPALTN---THHELRLDVEFKPLAP 276
 Db 1514 MGRGEFCER---VTEQDHTMPFLPEFNGF-SYLELNGLOTFLTCRQSMVEVFLAKSP 1569
 QY 277 DGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSLAVLSAEPLALGRWHRVSAERLN 336
 Db 1570 SGMIFYNGQKTDGKDFVSLALHDGYLEYRDLGKGAVALRSKEFVPLNTWISVLLERSG 1629
 QY 337 KDGSLRVNGGRVLRSSPGKSGQ-----LNLHTLLYLGGEVPSVPLSPATNNSAHFRGCV 391
 Db 1630 RKGVMRIINGERVMGESP-KSRKVPEHAFNLKEPFYVGGAPDFSKLARAALSTSFYNAV 1688
 QY 392 GEVSVNGKRLDITYFLSGGIGQCYQSDSPCRQPCQHGATCMPAGEYEFQCLCDRGKFG 451
 Db 1689 QRISIKGVPL-----LKEQHRSVAEIST-----FRA 1715
 QY 452 DLCEHEENPCQLREPCLGHTC---OQTRCLCLPCFSGPRCOQ-----GSGHGIA-- 498
 Db 1716 HCTOKPNECQ-----NGGTCSPLESYECACQGFSGAECVKVIEKAGDAEIAFD 1769
 QY 499 ---ESDWHLEGGGNDAPQGYAYFHDGFLAPGHVFSRSLPEVPETIELEVRTSTAGS 555
 Db 1770 GRTYMEYHNAVTKSPDA-----LDYPAEPSEKALQS--NHFELSITKEATQG 1814
 QY 556 LLLWQGVGEAGQCKDFISLGLQGHVFRYOLGSGEARLVSEDPINDGWHRRVTVLARE 615
 Db 1815 LILMS-----GKLESDVIALAVDGFQVMYDLGSKPVLRSVTPINTNTHIKAVRV 1870
 QY 616 GRGSIQVNGDELVSGRPGPNVAVNAKGSVYIGAP--DVATLTGGRFSSGIGTCVKNL 673
 Db 1871 QREGSLQVNGEAPITGSSPLGATQDLDGALWGLMERLSVAHKLPKAYSTGFIGCIRDV 1930

QY 674 VL 675
DB 1931 IV 1932

RESULT 5

T43060
agrin - electric ray (Discopyge ommata) (fragment)
C/Species: Discopyge ommata
C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 17-Nov-2000
C/Accession: T43060
R/Smith, M.A.; Magill-Solc, C.; Rupp, F.; Yao, Y.M.M.; Schilling, J.W.; Snow, P.; McMaha
submitted to the EMBL Data Library, September 1992
A/Reference number: Z22308
A/Accession: T43060
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: mRNA
A/Residues: 1-1328 <EMI>
A/Cross-references: EMBL:L01423; NID:G213102; PID:G213103; PIDN:AAA49224.1
C/Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repea
C/Keywords: glycoprotein; neuromuscular junction

Query Match 17.4%; Score 667; DB 2; Length 1328;
Best Local Similarity 27.0%; Pred. No. 6.2e-35;
Matches 200; Conservative 112; Mismatches 304; Indels 126; Gaps 24;

QY 2 IKITFRPDSADGMLLYNGQKRVPGSPNTLANRQPDIFSGLVGRPEFRFDAGSGWATIR 61
DB 673 ISMEFRASNDGLPLVQWTEKGG---LHFYRPS-----EGYVELRPNHGVNDGVIT 721
QY 62 HPTPLALGHHTVTLRLSLTQGLIIVGDLAPVNGTSGQKFGQLDNLNEEYLYGPDYD-- 119
DB 722 SKTLIRPGNHHVYVGNRRRSGMLSYDGEPLHIGESPCTDGLNLDLFLGCTPEDEMT 781
QY 120 -AIPKAGLSGFTGCVRELRIOCEEIVFHDLN-----LTAHGISHCPT--CRDRPCQNG 170
DB 782 LVTERITATKLGCGCRLDDVNN---LIYDLQERSNDVLYGSGVCGGNPCSPAPCKNR 838
QY 171 GQCHDESSYVVCVPAFTGSRG--EHSQALHCHPEACGPDPATCNRPDGRGYTCRCHL 228
DB 839 GKCHKEAEAFHESVGEFSGPTCADKHP--CDNPNCHQGANCVLPEG--GSKCECFM 894
QY 229 GRSLRCE-----SGVTVTTPSLGAGSYLALPAL-----TNTHELELDVEFFELAPD 277
DB 895 GREGEICERVSEAEQOGKAFIPEFNL--SYLEMNGIHFTVSDLLQKLSMEVIFLAKDPN 953
QY 278 GVLLFSGSGSPVEDFVSLAMVGHELFYELGSLGLAVLRSAEPALGRHWRVSAERLKN 337
DB 954 GMIFYNGQKTGGRGDFVSLNRDLRGYLEFFKYLKGAVALRSKAPIPLNWNVTVVERNGR 1013
QY 338 DGSIRVN-----GGRPVLRSSPGKSGQLNHTLLYLGVEPSPVLSPATNNSAHRGCV 391
DB 1014 KGLMKINDLVSGESPKSRKAP--HTALNLKEAFYVGGAPDNFKFAAAGIISGFTGAI 1071
QY 392 GEYSVNGKRLDLYTSFLSGIGQCVDSPCRQPCQHCATCMPAGEVEFFQCLCRDGFKG 451
DB 1072 QKLSLK-----SIFLLKENTENANEIS-----NFRW 1098
QY 452 DLCEHENPCQREPCLHGTC---QGTRCLCLPFGSGPRCQOGSGHGIAESDWHLEGS 507
DB 1099 HACTKTRNPCQ-----NGGVCSPRLREYDCMQRGFGFQCEKA-----LE-- 1139
QY 508 GGNDAPOQYCAVPHDGLAFPGHVFPSRLPVPETIELEVSTASGLLLMQGVVEGEA 567
DB 1140 -EKSAGSGSEVAFNGKTFEYHNVTVRSEKAVQVNFENSITEATKGLILMS-----GKI 1194
QY 568 GQCKDFISLQDGHILVFRYQLGSGEARLVSDPDINDGEHWRVTLALREGRGSIQVDEE 627
DB 1195 AEKSDYIALAVDGFQVMTYDLGSKPVTLRSTIPVNTQWVAIKANRIGHYGTQLQVNEA 1254
QY 628 LVSGRSPGNVAVNAGSVYIGAPDVALTGTGR-----FSSGITCVKVLVHSRPAAP 683
DB 1255 FVTGSSFFAATQDLDGALWLGIEKLA--PGNRLPKAYSTGTFIGCIKDWVIDR----- 1306

QY 684 PPQPLDLQHQRAQAGANTRPCFS 705
DB 1307 --QELQVEDALNNPTILHCEA 1326

RESULT 6

T19821
hypothetical protein ZC101.2e - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 23-Sep-2002
C/Accession: T19821; T19819; T19820; T27490; T27488; T27489; A47648; B47648; C47
R/Baynes, C.
submitted to the EMBL Data Library, March 1997
A/Reference number: Z19182
A/Accession: T19821
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-3375 <WIL>
A/Cross-references: EMBL:Z93375; PIDN:CAB07569.1; GSPDB:GN00020; CESP:ZC101.2e
A/Experimental source: clone C38C6
A/Accession: T19819
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-2441, 'R', 'KRKH', '3369', 'GN', '3372-3373', 'G', '3375', 'RLRHRRNAQNGPLSRKTRTTTKLFGSW
A/Cross-references: EMBL:Z93375; PIDN:CAB07567.1; GSPDB:GN00020; CESP:ZC101.2a
A/Experimental source: clone C38C6
A/Accession: T19820
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-1694, 'H', '1883-2441', 'R', 'KRKH', '3369', 'GN', '3372-3373', 'G', '3375', 'RLRHRRNAQNGPL
A/Cross-references: EMBL:Z93375; PIDN:CAB07568.1; GSPDB:GN00020; CESP:ZC101.2c
A/Experimental source: clone C38C6
R/Pericy, C.
submitted to the EMBL Data Library, March 1997
A/Reference number: Z20375
A/Accession: T27490
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-3375 <W12>
A/Cross-references: EMBL:Z93395; PIDN:CAB07708.1; GSPDB:GN00020; CESP:ZC101.2e
A/Experimental source: clone ZC101
A/Accession: T27488
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-2441, 'R', 'KRKH', '3369', 'GN', '3372-3373', 'G', '3375', 'RLRHRRNAQNGPLSRKTRTTTKLFGSW
A/Cross-references: EMBL:Z93395; PIDN:CAB07706.1; GSPDB:GN00020; CESP:ZC101.2a
A/Experimental source: clone ZC101
A/Accession: T27489
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-1694, 'H', '1883-2441', 'R', 'KRKH', '3369', 'GN', '3372-3373', 'G', '3375', 'RLRHRRNAQNGPL
A/Cross-references: EMBL:Z93395; PIDN:CAB07707.1; GSPDB:GN00020; CESP:ZC101.2c
A/Experimental source: clone ZC101
A/Accession: T27487
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-1128, 1290, 'DFARNSPS', '1299', 'NSS', '1303-1304', 'R', 'RHR', '1544-1545', 'RIRVRS', '155:
A/Cross-references: EMBL:Z93395; PIDN:CAB07704.1; GSPDB:GN00020; CESP:ZC101.2b
A/Experimental source: clone ZC101
R/Rogalaki, T.M.; Williams, B.D.; Mullen, G.P.; Moerman, D.G.
Genes Dev. 7, 1471-1484, 1993
A/Title: Products of the unc-52 gene in Caenorhabditis elegans are homologous to the cor
A/Reference number: A47648; MUID:9339574; PMID:8393416
A/Accession: A47648
A/Status: preliminary; nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-546, 'P', '548-2441', 'R', 'KRKH', '3369', 'GN', '3372-3373', 'G', '3375', 'RLRHRRNAQNGPLSR
A/Cross-references: GB:I13458
A/Accession: B47648
A/Status: preliminary; nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-546, 'P', '548-2198', 'D', '2290', 'NAR', '2294', 'L', '2296', 'WHATE', '2302-2303', 'V', '2305', 'I

1, 'ANIV', 2516-2517, 'LOGG', 2522, 'IDG', 2526, 'S', 2528, 'SRGFHV', 2535, 'F' <R02>
A;Cross-references: GB:L13458
A;Accession: C47648
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-546, 'P', 548-1128, 1290, 'D'ARNSPS', 1299, 'NSS', 1303-1304, 'R', 'RHR', 1544-1545,
A;Cross-references: GB:L13458
C;Genetics:
A;Gene: CESP:ZC101.2e; CESP:ZC101.2a; CESP:ZC101.2c; CESP:ZC101.2b
A;Map position: 2
A;Introns: 32/1, 134/1, 225/1, 335/2, 450/3, 739/3, 830/3, 860/2, 1064/2, 1129/1, 1158/3
2, 2613/1, 2684/1, 2757/1, 2813/3, 2863/1, 2900/3, 3094/1, 3176/1, 3250/2
C;Superfamily: LR11 protein; laminin-type EGF-like homology; LDL receptor ligand-binding
C;Keywords: extracellular matrix
F;149-183/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;190-224/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;233-268/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;553-1002/Domain: laminin-type EGF-like homology <LEG1>
F;1011-1058/Domain: laminin-type EGF-like homology <LEG2>

Query Match 13.4%; Score 511; DB 2; Length 3375;
Best Local Similarity 20.9%; Pred. No. 1.8e-24;
Matches 179; Conservative 140; Mismatches 332; Indels 204; Gaps 29;

QY 1 EKITRPSADGMLLYNGQKVP-----GSPTLNANRQPDFISFGLVGRPEFRPDAGS 55
DB 2557 DIEITLNTANPKGIIFET--KRINSGDLLATPYDTIHHEAKITDYGTV---LYEFIDGN 2610
QY 56 GWATTRHTPLALGHFHVTVLLRSITQGLSLVGLA-----PVNGTSQKPGQGLDLN 107
DB 2611 GRQIVETNPINPENNVIKNDKNQVTIQLNDESATIRHTNPLSLSTG-----VN 2664
QY 108 ELYLGGYDPDCAIPKAGLSGFTGCVRELIQOEIVFHDNLNTAHGISHCPTCRDRPC 167
DB 2665 RPFVIGGRHE-----PTNEANDFRGIISQVLSGHNVGLGARIPSSVVKYDACASTNL 2719
QY 168 QNGGCHDSES--SSVVCVPAGTGRCEHSAALCHPEACGPDPATCNVRDGRGVTCRC 226
DB 2720 LNCANCRNANHGFSCBAEFGVEYQW--KSNCHDESN--TGICLDNEE--SWQVCV 2775
QY 227 HLGRSLRCERGVVTTTP--SLSGAGSYLALPALNTNHELRLDVEFKPLAPDG---VLLF 282
DB 2776 PLGTTGLCEKEKTEIPQLGTSPTSFLAVRPVKFE---SIKMLRFOQADSDEHILMYF 2832
QY 283 SGGKSGPVEDFVSLAMVGHLEFFRYELGSLAVLRSAPALGRWHVSAERLNKQSLR 342
DB 2833 ASDVGSNTKQTSLSLIANQV-----LTVRRPKVEQKIRSETLEAGEL-IDVAVR 2883
QY 343 VNGRPVLRSSPGKSQGLNLHTL-----LYLGGVEPSVPLSPATNMSAHFRGCVGEVSVN 397
DB 2884 QAGNALVMTVDGNOVSTIETDLKPGTEIPGLPPGLN--SPDDVVSQSQGCVYELIN 2942
QY 398 GKRLDLYSFLSQG--IQCCYDS-----SPC 422
DB 2943 SQVDVL--QNLSSGDISSCESOFFVEEDTTTTTTEBEAVIEPTEETPTTEPI 3000
QY 423 ERQPCQHGATC-----MPAGEYEQCLCRDG-----448
DB 3001 TEEPEETPTTEETPTTEETPTTEETPTTEETPTTEETPTTEETPTTEETPTTEETPT 3057
QY 449 -----FKGDLCEHEN-----459
DB 3058 TSTTTTSTTEPEAPALVLPDTPVEENDVSDDEEISTISTVSPDNLGDSDSYSBGLT 3117
QY 460 -PCQLREPLHG-----GTQGTCLCLPGFSGPRCOQSGHGIAESDWHLEG--SG 508
DB 3118 PPSSEBEIVVDVYSTOEPNNICANSTC-----GMNGCQVPRNTHYTCCKLYDGPIC 3173
QY 509 GNDAPQGVAVHDDGFLAFPGHVFSTRSLPVPETILEVTRTASGLLLWQGVVEGAG 568
DB 3174 SLFKPIEHAARFDGDAFTLSLSDSEFPPLTSEKBEIVAFKFKTEQONGVLLWQ--QRPTVQ 3232
QY 569 QGKDFISLGLQDGLHVPFYQLSGSEARLVSPDINDGEWHRTALREGRGSIQVDGEEL 628

DB 3233 QMEDYISVGINVGHLSFSEYELGGAAHLISEERSVDDGKEHSVRFERKRGQMRIDNVR 3292
QY 629 VSGRSPGNVAVNAKGSVYIGAPDVATLTGSRFSGGITGCVKNLVLSHSPGAPPPQPL 688
DB 3293 VDRSGTGILAMLVNDGINFVGGVDPDISKATGSLFNNFVGCIADVELNGVK-----L 3344
QY 689 DLQHRAQAGANTRPC 703
DB 3345 DLNATAIDGKNVPC 3359

RESULT 7
C40228
neurexin II-alpha precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Jul-1995 #sequence revision 21-Jul-1995 #text_change 09-Dec-2002
C;Accession: C40228; S27886; S27887
R;Ushkaryov, Y.A.; Petrenko, A.G.; Geppert, M.; Suedhof, T.C.
Science 257, 50-56, 1992
A;Title: Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin recep
A;Reference number: A40228; MUID:92320296; PMID:1621094
A;Accession: C40228
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1715 <US>
A;Cross-references: GB:M96376; NID:G205714; PIDN:AAA41707.1; PID:G205716
A;Note: authors translated the codon GAC for residue 1464 as Thr and ACC for residue 146
R;Ushkaryov, Y.A.; Petrenko, A.G.; Geppert, M.; Suedhof, T.C.
submitted to the EMBL Data Library, July 1992
A;Description: Neurexins: Synaptic cell surface proteins related to the alpha-latrotoxin
A;Reference number: S27884
A;Accession: S27886
A;Molecule type: mRNA
A;Residues: 1-1666, 'CRK', 1670, 'PREKKLLPG', 1683-1685, 'GL', 1688, 'LDLA', 1694-1695, 'CCVCRCA
A;Cross-references: EMBL:M96376; NID:G205714; PIDN:AAA41706.1; PID:G205715
C;Genetics: 1666/2
A;Introns: 1666/2
C;Superfamily: neurexin; EGF homology
C;Keywords: alternative splicing; transmembrane protein
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-1715/Product: neurexin II-alpha #status predicted <MAT>
F;694-726/Domain: EGF homology <EGF>
F;1103-1135/Domain: EGF homology <EGF1>

Query Match 11.0%; Score 421; DB 2; Length 1715;
Best Local Similarity 21.6%; Pred. No. 5e-19;
Matches 193; Conservative 114; Mismatches 274; Indels 312; Gaps 37;

QY 2 IKITRPSADGMLLYN--GKRVPGSPNTLNANRQPDFISFGLVGRPEFRPDAGSGMATI 60
DB 517 ISLDFATTEPNGLLAFSQRRAGAGVGHSSSQRADYFAMELLDGLYLLLLDMSSGGIKL 576
QY 61 RHPT-PLALGHFHTVTLRSITQGLSLVGLA--PVNGTSQKFGQGLDNLNELYLGGYDPY 118
DB 577 RASSRKVNDGEWCHVDQFQDRGKSGISVNSRSTPPLAT--GESEVLDESELYLGLPEG 634
QY 119 GAIP-----KAGLSGFTGCVRELIQOEIVFHDNLNTAHGISHCP-----TC 162
DB 635 GRVDLPPLPEVWTAALRAGVCGVRDLFDGRSRLRLGLAEAGVAVPACSRETILKQC 694
QY 163 RDRPCQNGSQCHDSBSSVYVCV--PAGFTGSRCE-----HSQA 199
DB 695 ASAPCRNGGICREG--WNRFPVDCIGTGLGRVCREATVLSYDGSVMYMKMLPNAMHTEA 753
QY 200 -----LHCHPEACGPDATC-----213
DB 754 EDVSLRFMSQRAYGLMAMTTSRESADTLRLDLGGQMKLTWNLDCLRVGCCAPSAAGPE 813
QY 214 -----VNRP-----217
DB 814 TLFAGKLNDEWHVTLRVVRGKSLQLSDVNVTVVEGOMAGATRLLEFHNIEGTITERRF 873

QY 218 -----DGRGYTRCHLG-----RSLRCRGGVTTTPSLSGAGS 251
Db 874 ISVPSNFISGLGVNGPYDQCKGDITTYCELNARFGLBAIVADVPYFKRS---S 930
QY 252 YLALPALTNTTHHELRLDVEFKPLAPGVLLPFGSGKSGPVEDFVSLAWGHHLEFRYELGS 311
Db 931 YLALATL-QAYASMHLPFPKTTAPDGLLNFNSNG---NDFIVIELVKGYYHYVFDLGN 986
QY 312 GLAVLR--SAEPLALGRWHRSABR--LNKDGSLRVNGRVRPLRSSPKSGGLNLHTLLYL 368
Db 987 SPSLMKGNSDKPVNDQWNVVSRPGNVHTLKID-SRTVTOHNSG-RANLDLKGELYI 1044
QY 369 GVEPSVPLSPATNMSAH-----FRGCGEVSNGRLDLTYFSLGSGIGQCY-----DSS 420
Db 1045 GGLSKNN--FENLPKLVASRDGFOGLASVDLNGRLPLDLADAL--HRIGQVERGCDGPST 1101
QY 421 PCRQPCQHGATCWPAGEYEFQCLCRDGRFGKDLCEHENPQCLRFPCLHGGCQGRCLC 480
Db 1102 TCTEESCANQGVCLQW-----DGF-----TDCDT----- 1126
QY 481 LPQFSGPRCQ-----GSGHGIASDWHLEGGGNDAPGQYGFYHDDGFLAFPHVF 533
Db 1127 MTSYGGPVNDPGTYIFGKGALITVTM-----PNDREPSTR-----MDRLAVGFSTHOR 1177
QY 534 SRSLLPEVPEITIEVRTSTASGILLWQGVGEAGQGGKDFISLGLODGHLYFRYQLGSGE 593
Db 1178 SAVL-----VRVDSASGL-----GDYQLHIDQGTGVIFNVGTDD 1213
QY 594 ARLVSEDP-----INDGEHRTALREGRGSIQVD-----GEELVSGRSPGNVAV----- 640
Db 1214 --ITIDEPNAIVSDGKHVYFRTRSGGNATLQVDSWPVNERYPAGNFNRLAIAQRTP 1271
QY 641 -----NAKGSVYIGGADPVATLTCGRFSSGITCCVK 672
Db 1272 YRLGRVVDWLKGRQLTIFNSQAIIKIGRDQ-----GRPQGVSGLYN 1319
RESULT 8
T42218
slit-1 protein homolog - rat
N;Alternate names: MEGF4 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 16-Aug-2002
C;Accession: T42218
R;Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A;Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A;Reference number: Z14126; MUID:98360089; PMID:9693030
A;Accession: T42218
A;Status: preliminary; translated from GB/EMBL/DDJ
A;Molecule type: mRNA
A;Residues: 1-1531 <NAK>
A;Cross-references: EMBL:AB011530; NID:G3449289; PIDN:BA32460.1; PID:G3449290
A;Experimental source: strain Sprague-Dawley; Brain
C;Genetics:
C;Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein h

Query Match 11.0%; Score 419.5; DB 2; Length 1531;
Best Local Similarity 23.3%; Pred. No. 5.5e-19;
Matches 173; Conservative 78; Mismatches 245; Indels 245; Gaps 32;
QY 25 GSPTNLNANQDFISFLVGGR-PEPRF-----DAGSGMATIRHTPLALGH 70
Db 758 GIPKNVTELDGNQFTLVPGQLSTFKYLQVLDSNNKISSLSNFTNMSQLTLLISY 817
QY 71 FHTVTL-----LRSTQSLIVGLDAPVNGTSQKFGQLDNEELYLGVDYGAIPK 123
Db 818 NALQCIPLPAFQGLRSRLSLHNGD---VSTLOEGIFADVTSLSHLAIGANPLYCCHL 874
QY 124 AGLSS-----GFTGCGVRELRIOGEEIVFHDMLTAHG-----ISHCPT 161
Db 875 RNLSSWVKTYGKEGFIARCAGPPEMECK-----LLLTTPAKKFECCQGPSLAVQAKDP 928

QY 162 CRDRPCQNGGCHDSESSYVCYPAGFTGSRCEHSQALHCHPEACGPDATC-VNRPDGR 220
Db 929 CLSSPCQNGGCHNDPLVETRCTPSGKGRNCBVSILD-SCSSNPCCNGGTCHAEGEDA 987
QY 221 GYTCRCHLGRSLRCRGGVTTTPSLSGAGSYLALPALTNTTHHELRLDVEFKPLAPGVLL 280
Db 988 GFTCSGPFEGLTG-----DGIGNYTCQCPLOYTGRACEQLVDFCSFDLNPCCHEAQCV 1002
QY 281 LPSGKSGPVEDFVSLAWGHHLEFRYELGSLAVLSABEPLALGRWHRSABRLNKDGS 340
Db 1003 -----GNTTDDC 1009
QY 341 LR---VNGRPRVLRSSPKSQGLNLHT-----LLYLIG-GVEPSVPL-SPATNMSAHFRGCV 391
Db 1010 VKHDCWNGVCV-----DGIGNYTCQCPLOYTGRACEQLVDFCSFDLNPCCHEAQCV 1061
QY 392 GEVSNGHKLDTYSFLGSGIGQIC-YDSSPCERQPCQHGATCWPAGEYEFQCLCRDGF 450
Db 1062 G--TPEGPRCEVCPGYTGD-----NCSKNQDDCKDHQCCQGAQCCVDEIN-STACLCAGYS 1114
QY 451 GDLCE--HEENPQCLREPCLHGGTC--QGTR--CLCLPGFSGPRCQCGSGHGAESDWH 503
Db 1115 GOLCEIIPAPRNSCEGTE-CQGANCVDDQSGRPVCQCLPGFGGPECEKLLSVNFPVDRDTY 1173
QY 504 LEGSGNDAPQCYGAYFHDGFLAPGHVFSRSLPEVPE-TIELEVRTSTASGILLWQGV 562
Db 1174 LQ-----FTD-----LQNWPRANITLQVSTAEQDNGHLLYNG- 1204
QY 563 EVGAGQKQDFISLGLODGHLYFRYQLGS-GEARLVSEDPINDGEHRTALREGRGSI 621
Db 1205 -----DNDHIAVELYQGHVRSYDPSYPSAIYSATINDQGFHTVELVTFDQMVNL 1257
QY 622 QVDGEELVSGRSPGNVAVNAKGSVYIGGAP-DVAT-----LTGGRFSSGITCCVK 672
Db 1258 SIDGSPMTMDNFGKHVTLNSEAPLYVGMFVDVNSAAFLWQLNGTSPH----GCIRN 1313
QY 673 LVLES-----ARPGAPP 684
Db 1314 LYINNELQDFTKQMKGVVP 1334
RESULT 9
I45944
neurexin I-alpha - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 19-Dec-1997 #sequence_revision 19-Dec-1997 #text_change 09-Dec-2002
C;Accession: I45944
R;Ullrich, B.; Ushkaryov, Y.A.; Sudhof, T.C.
Neuron 14, 497-507, 1995
A;Title: Cartography of neurexins: more than 1000 isoforms generated by alternative splicing
A;Reference number: I45944; MUID:95209856; PMID:7695896
A;Accession: I45944
A;Status: preliminary; translated from GB/EMBL/DDJ
A;Molecule type: mRNA
A;Residues: 1-1530 <ULL>
A;Cross-references: GB:LI44855; NID:G388560; PIDN:AAA74123.1; PID:G388561
A;Superfamily: neurexin; EGF homology
F;703-735/Domain: EGF homology <EGF>
Query Match 10.6%; Score 405.5; DB 2; Length 1530;
Best Local Similarity 20.5%; Pred. No. 4.4e-18;
Matches 203; Conservative 112; Mismatches 285; Indels 391; Gaps 42;
QY 2 IKITPRDSDAGMLIY-NGQKRVFGSPNTNLNANQDFISFLVGGRPEPRFDAGSGMATI 60
Db 527 ISDFPRTEPNGLILPSHGKPRHOKDAKHPQMKVDFFAIEMLDGHLVLLDMSGGTIKI 586
QY 61 RHPTPLAL-----GHFTVTLRLSLTQSLIVGLD-APVNGTSQKFGQLDNEELYLG 113
Db 587 K-----ALQKKNVDEGWTHVDFQDRSGTISVNTLTPY--TAPGSQLDLDDELYLG 639
QY 114 GYPDYGAIPKAG-----LSSGFIGVRELRIOGEEIVFH---DLNLTAHGISHC 159

Db 640 GLPE-----NKAGLVPTTEVNTALLNYGVGIRDLFDGQSKDIQMAEQVSTAGVKPSC 695
QY 160 -----PTCRDRPCQNGGCHDSSESSVVCVCP-AGFTGSRCE----- 195
Db 696 SRETAQPCLSNPKNGMCRDG-WNRVYVDCSGTGLGRSCEREATVLSYDGSFMFKIQL 754
QY 196 -----HSQA-----LHCHPEAC-- 207
Db 755 PVMHTTEADSVLSRFRSQRAYGILMATTSDTLRLDLAGRVKLTVNLDCIRINCNS 814
QY 208 --GPDATCV--NRPDGRGYTCR----- 225
Db 815 SKGPETLFAGYNLDNENWHTVVRVGRGSKLTVDDQQAQMGAGDHTLRFHNIETGI 874
QY 226 -----CHL-GRSGLR--CEBQVTVTT 243
Db 875 ITERRYLSSVPSNFIQHLQSLTFNGMAYIDLCKNGDIDYCELNARFGFRNIADPVTFKT 934
QY 244 PSLGAGSYLALPALTNTTHELRDLVEFKPLADPGVLLFSGKSGPVEDFVSLAMVGGHL 303
Db 935 KS-----SYVALATL-QAYTSMHLFPQFTKTSLDGLILYN--SGDGNDFIVVELVKGYL 985
QY 304 EFRYELGSLAVLR--SAEPALGRHRYSAER-----LNKDGSLRVNGRPPVLS 352
Db 986 HYVFDLGNANLKGSSNKPLNDNQHNVMISRDTSNLHTVKIDTKITQITAG----- 1039
QY 353 SPKGQGLNHLTLVLLGGV--EPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLYSFLGS 410
Db 1040 -----ARNLDKSLDIYGGVAKETYSKPLVHAKEGFGQCLASVDLNGRLPDL----- 1088
QY 411 QGIGQCYDSSPCERQPCQHGATCMPAGEYFQCLCRDGFKGLDCEHENPCQLREPCPLHG 470
Db 1089 -----ISDALFCN-----GQIE-----RG--CEGPTTTCQ--EDSCSNQ 1118
QY 471 GTC-----QGTCLC-LPGFSGPRCQ-----GSGHGIAESDWHLESGGNDAPQGYGA 518
Db 1119 GVCLOQWDGISCDCSMTSFGPLCNDPGTTFYFSKGGQITVKW-----PNDRPSTRA- 1172
QY 519 YFHDGDLFAPGHVFRSRLPEVPETIELE---VRTSTASGLLLMQGVEVSGAGQKOFIS 575
Db 1173 ---DRLAIGF-----STVQKEAVLRVDSSSL-----GDYLE 1202
QY 576 LGLQDGHVFRYOLGSGEARLVSEDP--INDGEWHRVTLRREGSRGSIQVDG---BELVS 630
Db 1203 LHHQKIGKVFNGVTDDIAESNAIINDGKYHVVRFRSGGNATLQVDSWPVIERYPA 1262
QY 631 GRSPGPNVAV-----NAKGSVYICG----- 650
Db 1263 GNNDRERLAIAQRIPYRLGRVVDWMLDKGRLTFNSQATIIIGKQEQGPPFQQLSG 1322
QY 651 -----APDVALTGGRFSSGITGCVKXLY 674
Db 1323 LYTNGLKVLNMAAENDANIAIVGNVRLVGEVSPSSMTTESTATAMQSEMSTIMETTTTLA 1382
QY 675 LHGARGAPPPPOPLDQHQRAQAGANTRPCPS 705
Db 1383 TSTARCKPTKEPVSQTTDDILVASAECPS 1413

RESULT 10

A40228
A;neurexin I-alpha precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Dec-2002
R;Ushkaryov, Y.A.; Petrenko, A.G.; Geppert, M.; Suedhof, T.C.
Science 257, 50-56, 1992
A;Title: Neurexins: synaptic cell surface proteins related to the alpha-latrogenin recep
A;Reference number: A40228; PMID:92320296; PMID:1621094
A;Accession: A40228
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA

A;Residues: 1-1507 <US>
A;Cross-references: GB:M96374; NID:G205710; PIDN:AAA41704.1; PID:G205711
C;Superfamily: neurexin; EGF homology
C;Keywords: alternative splicing; transmembrane protein
F;1-30/Domain: signal sequence #status predicted <Sig>
F;1-1507/Product: neurexin I-alpha #status predicted <Mat>
F;680-712/Domain: EGF homology <EGF>
F;1087-1119/Domain: EGF homology <EGF1>

Query Match 10.5%; Score 403.5; DB 2; Length 1507;
Best Local Similarity 20.4%; Pred. No. 5.8e-18;
Matches 202; Conservative 113; Mismatches 285; Indels 391; Gaps 42;

QY 2 IKITRPDSADQMILY-NGCKRVPGSPNTLANRQDFISFGLVGRPRFRFDAGSGMATI 60
Db 504 ISFDFTTTPNGLLIFESHGKPRHQDAKHPQMKVDFFAIEMLDGHLVLLDMSGSTIKI 563
QY 61 RIHTPAL-----GHFTVTLRLSLTQSLIVGLD-APVNGTSQKQGLDLNLELYLG 113
Db 564 K-----ALQKYNDEGWYHVPQDRSGTISVNTLTPY--TAPGESEILDDELDELGL 616
QY 114 GVPDYGAIPKAG-----LSGFGICVRELRIQGEIIVFH---DLNLTAGHSHC 159
Db 617 GLPE---NKAGLVPTTEVNTALLNYGVGIRDLFDGQSKDIQMAEQVSTAGVKPSC 672
QY 160 -----PTCRDRPCQNGGCHDSSESSVVCVCP-AGFTGSRCE----- 195
Db 673 SRETAQPCLSNPKNGMCRDG-WNRVYVDCSGTGLGRSCEREATVLSYDGSFMFKIQL 731
QY 196 -----HSQA-----LHCHPEAC-- 207
Db 732 PVMHTTEADSVLSRFRSQRAYGILMATTSDTLRLDLAGRVKLTVNLDCIRINCNS 791
QY 208 --GPDATCV--NRPDGRGYTCR----- 225
Db 792 SKGPETLFAGYNLDNENWHTVVRVGRGSKLTVDDQQAQMGAGDHTLRFHNIETGI 951
QY 226 -----CHL-GRSGLR--CEBQVTVTT 243
Db 852 ITERRYLSSVPSNFIQHLQSLTFNGMAYIDLCKNGDIDYCELNARFGFRNIADPVTFKT 911
QY 244 PSLGAGSYLALPALTNTTHELRDLVEFKPLADPGVLLFSGKSGPVEDFVSLAMVGGHL 303
Db 912 KS-----SYVALATL-QAYTSMHLFPQFTKTSLDGLILYN--SGDGNDFIVVELVKGYL 962
QY 304 EFRYELGSLAVLR--SAEPALGRHRYSAER-----LNKDGSLRVNGRPPVLS 352
Db 963 HYVFDLGNANLKGSSNKPLNDNQHNVMISRDTSNLHTVKIDTKITQITAG----- 1016
QY 353 SPKGQGLNHLTLVLLGGV--EPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLYSFLGS 410
Db 1017 -----ARNLDKSLDIYGGVAKETYSKPLVHAKEGFGQCLASVDLNGRLPDL----- 1065
QY 411 QGIGQCYDSSPCERQPCQHGATCMPAGEYFQCLCRDGFKGLDCEHENPCQLREPCPLHG 470
Db 1066 -----ISDALFCN-----GQIE-----RG--CEGPTTTCQ--EDSCSNQ 1095
QY 471 GTC-----QGTCLC-LPGFSGPRCQ-----GSGHGIAESDWHLESGGNDAPQGYGA 518
Db 1096 GVCLOQWDGISCDCSMTSFGPLCNDPGTTFYFSKGGQITVKW-----PNDRPSTRA- 1149
QY 519 YFHDGDLFAPGHVFRSRLPEVPETIELE---VRTSTASGLLLMQGVEVSGAGQKOFIS 575
Db 1150 ---DRLAIGF-----STVQKEAVLRVDSSSL-----GDYLE 1179
QY 576 LGLQDGHVFRYOLGSGEARLVSEDP--INDGEWHRVTLRREGSRGSIQVDG---BELVS 630
Db 1180 LHHQKIGKVFNGVTDDIAESNAIINDGKYHVVRFRSGGNATLQVDSWPVIERYPA 1239
QY 631 GRSPGPNVAV-----NAKGSVYICG----- 650
Db 1240 GNNDRERLAIAQRIPYRLGRVVDWMLDKGRLTFNSQATIIIGKQEQGPPFQQLSG 1299

QY 651 -----APDVATLTGGRFFSGITGCVKNLV 674
 Db 1300 LYTNGLKVLNMAENDANIALVGNRLVGEVSPMTTESTATAMQSEMSTIMETTTTLA 1359
 QY 675 LHSARPQAPPQPLDLQRAQAAGANTPCPS 705
 Db 1360 TSTARRGKPTKERTISQTTDDILVASAECPS 1390
 RESULT 11
 T42626
 secreted leucine-rich repeat-containing protein SLR12 - mouse (fragment)
 N:Alternate names: neurogenic extracellular slit protein
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 16-Aug-2002
 C:Accession: T42626
 R:Holmes, G.P.; Negus, K.; Burridge, L.; Raman, S.; Algar, E.; Yamada, T.; Little, M.H.
 Mech. Dev. 79, 57-72, 1998
 A:Title: Distinct but overlapping expression patterns of two vertebrate slit homologs in
 A:Reference number: Z22177; MUID:99279238; PMID:10349621
 A:Accession: T42626
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1025 <HOL>
 A:Cross-references: EMBL:AF074960; NID:94151258; PID:94151259; PIDN:AAD04345.1
 C:Genetics:
 A:Gene: Slit2
 C:Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein
 Query Match 10.1%; Score 387.5; DB 2; Length 1025;
 Best Local Similarity 29.7%; Pred. No. 4e-17;
 Matches 108; Conservative 55; Mismatches 144; Indels 57; Gaps 18;
 QY 158 HC-----PTCRDRPCQNGGQCHDSSESYVVCVPCAGFTGSRCEHSQAL-----HCHPEAC 207
 Db 568 HCDIDFDDCDNCKNGAHCCTDA-VNGYTCVCEGYSGLFCFSPZPMLPRTSPCDNFDC 626
 QY 208 GPDATC---VNRPDGRVTCRCHLGRSLRCEEGVTTPSLGAGSYLALPAL-----T 259
 Db 627 QNGAQCIIRNEP-----IOCLPGYIGKECKLVSV---NFNKGSIYLPQIPAKVRPQT 678
 QY 260 NTHHELRLDVEFKPLAPDGVLLFSGGSGPVEDFVSLAMVGGHLEFRYELGSLA-VLRAS 318
 Db 679 NITLQIATDED-----SGILLYKGDK-----DHIAVELYRGRVRSYDTGSHPASAIYS 727
 QY 319 AEPLALGWHVRSALRNKQSLRVNGRVPVLRSSPKSQGLNLTLLYLGVE-----P 373
 Db 728 VETINDGNFHIVELLTLDLSLSLVSDGSGPKVIITNLSKQSTLNFDPFLYVGGMPGKNVVA 787
 QY 374 SVPLSPATNMSAHPGCVGEVSVNGKRLDLYTSLGSGQIGCCQVDSGPCRQPCQHGATC 433
 Db 788 SLRQAPQGN-GTSPHGCIRNLYINSELQDFKPMQGIPLPGC---EPCHKVCAHG-MC 842
 QY 434 MPAGEYFQCLCRDGFYKGDICEHEN-PCQLREPCLLGGTCQ-----GTRCLCLPFGSGP 487
 Db 843 QPSSQSQFTCECEGNGWGPLCDQRTDQDPC-LGNKCVH-GTCLPINAFSYSCKLEGHGGV 900
 QY 488 RCOQ 491
 Db 901 LCDE 904
 RESULT 12
 T13953
 MEGF5 protein - rat
 N:Alternate names: slit protein homolog
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 16-Aug-2002
 R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
 Genomics 51, 27-34, 1998
 A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
 A:Reference number: Z14126; MUID:98360089; PMID:9693030

A:Accession: T13953
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1523 <NAX>
 A:Cross-references: EMBL:AB011531; NID:g3449291; PIDN:BA032461.1; PID:g3449292
 C:Genetics:
 A:Gene: MEGF5
 C:Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein
 Query Match 10.1%; Score 387; DB 2; Length 1523;
 Best Local Similarity 28.6%; Pred. No. 6.7e-17;
 Matches 113; Conservative 52; Mismatches 148; Indels 82; Gaps 18;
 QY 159 CPT-----CRDRPCQNGGQCHDSSESYVVCVPCAGFTGSRCEHS-----QALLHCHPEACG 208
 Db 1071 CETDNDCCVAHKRGAQCQVDA-VNGYTCICPQGFSGLFCEHPPMVLVLTSPCDQVECC 1129
 QY 209 PDATCV---NRPDGRVTCRCHLGRSLRCEEGVTTPSLGAGSYLALPALTNHHHEL 265
 Db 1130 NGAQCIVVQOEP-----TCRCPPGAPGPRCEKLTIV---NFVKGDSYVEL-ASAKVRPQA 1180
 QY 266 RLDFVEFKPLAPDGVLLFSGGSGPVEDFVSLAMVGGHLEFRYE-LGSLGLVLRSAEPLAL 324
 Db 1181 NISLQVATDKONGILLYKGD-----NDPLALELQGHVRLVYDLSLSPPTTVYSVETVND 1235
 QY 325 GRWHRVSAERLNKQSLRVNGRVPVLRSSPKSQGLNLTLLYLGVEPSPVLPSP-----A 380
 Db 1236 GQFHSVELVMLNQTLLNLYVDKAPKSLGKLOKQPAVGINSPYLGLGPISTGLSALROGA 1295
 QY 381 TNMSAHERGCVGEVSVNGKRLDLYTSLGSGQIG-----CLCLPGESGPRCCQ 491
 Db 1296 DRPLGGFHGCIHEVRINNELQD--FKALPPQSLGVSPCKSCTVCRHGLCRSVKDSVVC 1353
 QY 415 QCYD-----SSPCERQPCQHGATCMPAGEYFQCLCRDGFYKGDICEHEE---NP 460
 Db 1354 ECHPGWTGPLCDQEAQDFCLGHSCSHG-TCVATGN-SYVCKAEGYEGPLCDQKNDNSANA 1411
 QY 461 COLREPCLLGGTCQGTR-----CLCLPGESGPRCCQ 491
 Db 1412 CSAFK-CHHGQCHISDRGEPCYCLCPGFGSNHCEQ 1445
 RESULT 13
 IJFPTM
 cadherin-related tumor suppressor precursor - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Feb-1997
 C:Accession: A41087; B41087
 R:Mahoney, P.A.; Weber, U.; Onofrechuk, P.; Biesemann, H.; Bryant, P.J.; Goodman, C.S.
 Cell 67, 853-868, 1991
 A:Title: The fat tumor suppressor gene in Drosophila encodes a novel member of the cadhe:
 A:Reference number: A41087; MUID:92069752; PMID:1959133
 A:Accession: A41087
 A:Molecule type: mRNA
 A:Residues: 143-485; 1279-5147 <NAH>
 A:Cross-references: GB:M80537
 A:Accession: B41087
 A:Molecule type: DNA
 A:Residues: 1-142; 487-1278 <MA2>
 A:Cross-references: GB:M80537
 A:Note: 1229-Gly and 1233-Ser were also found
 C:Genetics:
 A:Gene: fat
 A:Cross-references: FlyBase:FBgn0001075
 C:Superfamily: cadherin-related tumor suppressor; cadherin repeat homology; EGF homology
 C:Keywords: calcium binding; cell adhesion; duplication; transmembrane protein
 F:1-35/Domain: signal sequence #status predicted <SIG>
 F:36-5147/Product: cadherin-related tumor suppressor #status predicted <MAT>
 F:36-5147/Domain: extracellular #status predicted <EXT>
 F:51-156/Domain: cadherin repeat homology <CR1>
 F:159-270/Domain: cadherin repeat homology <CR2>
 F:271-382/Domain: cadherin repeat homology <CR3>
 F:390-494/Domain: cadherin repeat homology <CR4>

F.497-599/Domain: cadherin repeat homology <CR5>
F.602-708/Domain: cadherin repeat homology <CR6>
F.718-822/Domain: cadherin repeat homology <CR7>
F.831-942/Domain: cadherin repeat homology <CR8>
F.948-1049/Domain: cadherin repeat homology <CR9>
F.1052-1153/Domain: cadherin repeat homology <CR10>
F.1156-1278/Domain: cadherin repeat homology <CR11>
F.1281-1384/Domain: cadherin repeat homology <CR12>
F.1387-1489/Domain: cadherin repeat homology <CR13>
F.1492-1601/Domain: cadherin repeat homology <CR14>
F.1607-1713/Domain: cadherin repeat homology <CR15>
F.1717-1823/Domain: cadherin repeat homology <CR16>
F.1826-1922/Domain: cadherin repeat homology <CR17>
F.1925-2027/Domain: cadherin repeat homology <CR18>
F.2028-2167/Domain: cadherin repeat homology <CR19>
F.2169-2278/Domain: cadherin repeat homology <CR20>
F.2281-2384/Domain: cadherin repeat homology <CR21>
F.2387-2491/Domain: cadherin repeat homology <CR22>
F.2494-2596/Domain: cadherin repeat homology <CR23>
F.2599-2703/Domain: cadherin repeat homology <CR24>
F.2707-2810/Domain: cadherin repeat homology <CR25>
F.2813-2913/Domain: cadherin repeat homology <CR26>
F.2915-3013/Domain: cadherin repeat homology <CR27>
F.3014-3124/Domain: cadherin repeat homology <CR28>
F.3127-3229/Domain: cadherin repeat homology <CR29>
F.3232-3334/Domain: cadherin repeat homology <CR30>
F.3337-3439/Domain: cadherin repeat homology <CR31>
F.3442-3545/Domain: cadherin repeat homology <CR32>
F.3548-3651/Domain: cadherin repeat homology <CR33>
F.3654-3756/Domain: cadherin repeat homology <CR34>
F.3759-3851/Domain: EGF homology <EG1>
F.3954-4010/Domain: EGF homology <EG2>
F.4017-4048/Domain: EGF homology <EG3>
F.4056-4089/Domain: EGF homology <EG4>
F.4096-4127/Domain: EGF homology <EG5>
F.4584-4609/Domain: transmembrane status predicted <TM>
F.4610-5147/Domain: intracellular status predicted <INT>

Query Match 10.0%; Score 383; DB 1; Length 5147;
Best Local Similarity 12.9%; Pred. NO. 4.7e-16;
Matches 185; Conservative 97; Mismatches 265; Indels 260; Gaps 37;

QY 4 ITRPDSADGMLLYNGKRVPCSPNLL-----ANRQDFISFLVGGPERED 52
DB 3883 IKSRSNDEVLVYLLSGGSGNSTLQMLMAVLAKTSYQQPYLIERLEKESAF--- 3939

QY 53 AGSGMATIRHPTPLALGHFHTVTLRLSTQGLVIG-----DLAPVNGTSQKFGOLD 105
DB 3940 -----SELLQKEVIVGVEPCPEPDVCENGVCVSATMLLD 3974

QY 106 LNE-----ELYLOG---YPPYGAIPKAGLSSGFIG---CYR-----ELRI 139
DB 3975 AHSFVIQSPALVSGPRVVDY-----SCQCTSGFSGQCRRODCLPNPCHSOVQCR 4030

QY 140 QGEETVFDLNLTAHGISHC-----PTCRDRPCQNGGQCHDS-ESSSYVVCVPAGFTGSR 193
DB 4031 LGSD---FQCMCPANRDKHCKERSDVCKSPCRNGGSCQSPDGSSYFCLCRPFRGNQ 4088

QY 194 CHEQALHCHPACGPDATCUN-RPDGRTYCRHLGSLRCBEGVTVTTPSLSGAGSY 252
DB 4089 CB-SVSDSCRPNCLHGLGCLVSLKP---GYKNCNTPGRYRHCERFSYGFQPL-----SY 4139

QY 253 LALPALTNTHRELVDVEPKLAPDGVLLFS-GKSGSPVDFVSLAMVGVGHLEFRYELG- 310
DB 4140 MTFPALDVTND--TSIVPATTKPSLLLYNYGMQSGRSDFLAIELV--HDAVFSSGG 4195

QY 311 --SGLAVLRASAPLALGRVHRVSRNLKDGSLRV-----NGRPVLRSSPKSQ----- 358
DB 4196 ARTAISTVIAGRLADGGHKVKTATNRGRVMSLSVAKCADSDGVCTECLPGDSSCIYADEV 4255

QY 359 ----GLNLHTL-LYLGGVPSVPL--SPATNSAHRGCVGEVSNVGRKLDLTYSFLGSG 411
DB 4256 GPVGTILNFKNQPMITGGISSADPILRPFQVQSHSDVLVCLHSVHLGGRLNLS-SPLOQK 4314

QY 412 GIGQCYDSSPCRPQCOHGATCMPAGEYEFQCLCRDGFKGLCEHENPCQLRBPCLHGG 471
DB 4315 GI-----LAGCNRQACQ-----PALAAE 4332

QY 472 TCGQTRCLCPGFSRQCGSGHGIABSDWHLESGGND--PGQYGAIFHDDGFLAAPP 529
DB 4333 RCGFAGQCIDRWSSSLQCQGG-----HLQSPDCSDSLEPITLG---EGAFVEFR 4379

QY 530 -GHVFSR-----SLPEVPETIELEVRTSTASG 555
DB 4380 ISEIYRMQLLDNLNYSKASWLDNQMRRAVNFSTASQIYEAPKMLSLFRYKQGG 4439

QY 556 LLLMQGVEVGEAGQKDFSLGLQDHLVRYQLGSGEARLVSE-DPINDGEWHRVTALR 614
DB 4440 QILY-----RATQMFSLSLRGLVYYSKQHLTINMTVQETSLNDGKHNVSLES 4492

QY 615 EGRRSIQVDGSELVSGRSPGNVAVNAKGSVYIGADPDVATL-----TGRGF 662
DB 4493 EGRSLRLIYDGRQ-----VGDELDIAGVHDFLDPLYLILANVGG-- 4530

QY 663 SSGITGCVKVLVHSGARPGAPPPQLD 689
DB 4531 -EAFVGLANVTNNEL-----QPLN 4550

RESULT 14
A48216
neurexin III-alpha secreted type 1 precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 03-Dec-2002
C;Accession: A48216; B48216
R;Ushkaryov, Y.A.; Suedhof, T.C.
A;Title: Neurexin IIIalpha: extensive alternative splicing generates membrane-bound and
A;Reference number: A48216; MUID:93342001; PMID:8341647
A;Accession: A48216
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1438 <USH>
A;Cross-references: GB:L14851
A;Accession: B48216
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1368,1372-1438 <US2>
A;Cross-references: GB:L14851
C;Genetics:
A;Introns: 1372/1
C;Superfamily: neurexin; EGF homology
C;Keywords: alternative splicing; brain; cell surface component; duplication; extracellu
F;1-27/Domain: signal sequence #status predicted <SIG>
F;202-234/Domain: EGF homology <EGF>
F;651-683/Domain: EGF homology <EGF1>

Query Match 10.0%; Score 382; DB 2; Length 1438;
Best Local Similarity 19.8%; Pred. NO. 1.3e-16;
Matches 175; Conservative 138; Mismatches 273; Indels 296; Gaps 38;

QY 1 EIKITRPSADGMLLYNGKRVPCSPNLANRQDPDFISFLVGGPEREPFDAGSG-MAT 59
DB 286 EITLSEKTPWRNGLIHTG-----KSDYVNLAKDGNVSLVNLGSAFPA 332

QY 60 IRHPTFLALGHP-----HTVTLRLSTQGLVIGDLAPVNGTSQKFGOLDNEELYGG 114
DB 333 IVEPVN---GKENDNAWHDVKVTNLRQVTISVDGILTTTGYTQEDYTMGSDDFYVGG 389

QY 115 YPDYGAIPKAGLSSGFIGVCRHLRIGEEIVFDNLNTAHGISHCPTCRDRPCQNGQCH 174
DB 390 SPETADLPQSPVSNPMGCKEVVYKNDIRLESLRARI----- 430

QY 175 DSSSSYVVCVCPAGFTGSCRSQALHCHPEACGPDATCVRNPDGRGYTCRCHLGRSLR 234
DB 431 DTRMKIY-----GEVVEK 443

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:19:08 ; Search time 51.7431 Seconds
(without alignments)
2876.963 Million cell updates/sec

Title: US-10-006-011a-3

Perfect score: 3825

Sequence: 1 EIKITRPDSADGMLLYNGQ.....QPLDLQRAQAQANTPCPS 705

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	901.5	23.6	1931	15	US-10-094-886-180
2	862.5	22.5	1940	13	US-10-016-283-34
3	840	22.0	775	15	US-10-108-260A-4433
4	832	21.8	819	10	US-09-866-050A-503
5	640.5	16.7	492	13	US-10-016-283-36
6	628	16.4	463	15	US-10-104-047-3058
7	605	15.8	432	9	US-09-764-853-541
8	605	15.8	432	9	US-09-764-898-179
9	605	15.8	432	10	US-09-764-881-102
10	605	15.8	432	14	US-10-073-865-78
11	605	15.8	432	15	US-10-242-747-102
12	579.5	15.2	456	13	US-10-016-283-25
13	576	15.1	440	13	US-10-016-283-26
14	543.5	14.2	375	9	US-09-978-249-10
15	528	13.8	390	13	US-10-016-283-27

16 441 11.5 3313 9 US-09-737-149-29
17 441 11.5 3313 16 US-10-038-854-69
18 420.5 11.0 1712 15 US-10-291-172-359
19 419.5 11.0 1531 11 US-09-970-944-31
20 412.5 10.8 1531 11 US-09-970-944-28
21 407 10.6 1534 11 US-09-970-944-30
22 403 10.5 3301 15 US-10-038-854-68
23 398 10.4 1438 15 US-10-287-839-2
24 397.5 10.4 1525 14 US-10-289-776-2
25 397.5 10.4 1529 10 US-09-766-511B-29
26 397.5 10.4 1529 14 US-10-189-123-67
27 397.5 10.4 1529 14 US-10-188-495-67
28 397.5 10.4 3312 14 US-10-225-567A-656
29 397.5 10.4 3312 16 US-10-038-854-67
30 397.5 10.4 4115 16 US-10-038-854-4
31 394.5 10.3 1523 10 US-09-946-374-198
32 394.5 10.3 1523 11 US-09-970-944-29
33 394.5 10.3 1523 13 US-10-052-586-290
34 394.5 10.3 1523 14 US-10-174-590-290
35 394.5 10.3 1523 14 US-10-176-758-290
36 394.5 10.3 1523 14 US-10-175-737-290
37 394.5 10.3 1523 14 US-10-173-706-290
38 394.5 10.3 1523 14 US-10-175-738-290
39 394.5 10.3 1523 14 US-10-175-752-290
40 394.5 10.3 1523 14 US-10-176-482-290
41 394.5 10.3 1523 14 US-10-176-757-290
42 394.5 10.3 1523 14 US-10-176-913-290
43 394.5 10.3 1523 14 US-10-180-552-290
44 394.5 10.3 1523 14 US-10-180-557-290
45 394.5 10.3 1523 14 US-10-173-700-290

ALIGNMENTS

RESULT 1
US-10-094-886-180
; Sequence 180, Application US/10094886
; Publication No. US20040002120A1

GENERAL INFORMATION:

APPLICANT: Kekuda, Ramesh
APPLICANT: Tchernev, Velizar T.
APPLICANT: Liu, Xiaohong
APPLICANT: Spytek, Kimberly A.
APPLICANT: Patturajan, Meera
APPLICANT: Burgess, Catherine
APPLICANT: Vernet, Corine A.
APPLICANT: Li, Li
APPLICANT: Gorman, Linda
APPLICANT: Malyankar, Uriel M.
APPLICANT: Boidog, Ferenc
APPLICANT: Guo, Xiaojia
APPLICANT: Shenoy, Suresh
APPLICANT: Padigaru, Muralidhara
APPLICANT: Taupier, Raymond J., Jr.
APPLICANT: Miller, Charles
APPLICANT: Casman, Stacie
APPLICANT: Pena, Carol
APPLICANT: Gangolli, Esha
APPLICANT: Gusev, Vladimir
APPLICANT: Smithson, Glennnda
APPLICANT: Zerhusen, Bryan
APPLICANT: Gerlach, Valerie
APPLICANT: Pochart, Pascal
APPLICANT: Fernandes, Elma
APPLICANT: Shinkets, Richard
APPLICANT: Rastelli, Luca
APPLICANT: Spaderna, Steven
APPLICANT: Laroche, William
APPLICANT: Zhong, Mei

; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-290 B
; CURRENT APPLICATION NUMBER: US/10/094,886

; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/274,322
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,182
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/288,052
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/318,510
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/314,018
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/296,693
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/313,626
; PRIOR FILING DATE: 2001-08-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 298
; SOFTWARE: Patent in 2.1
; SEQ ID NO 180
; LENGTH: 1931
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-886-180

Query Match 23.6%; Score 901.5; DB 15; Length 1931;
Best Local Similarity 32.2%; Pred. No. 1.7e-64;
Matches 242; Conservative 93; Mismatches 266; Indels 151; Gaps 22;

QY 2 IKITRPSADGMILYKQKRVPGSPPTNLARQDFTSFGLVGGRPEFRFDAGSGMATIR 61
Db 1282 LALFRALEPOLLLYNGNA-----RGKDFLALALDDGRVQTRDFTSGSPAVLT 1330
QY 62 HPTPLALGHFTVTLRLSLTQGLSIVGDLAPVNGTSQKFGQGLDNEELYLGVPD--Y 118
Db 1331 SAVFVPEQWHRLELSRHRWGTLSDVGETPVLGESPGTDLNLDLTFVGGVPEQAA 1390
QY 119 GAIPKAGLSGFGICVRELRTQGEIVFHDNLN-----TAHGISHC--PTCRDRPCQNG 170
Db 1391 VALERTFVAGLRGICRILLDNNQRL---ELGIFGAATRGSGVGECDHPCLNPCHGG 1447
QY 171 QCHDSESSYVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
Db 1448 APCQNLAEGRFHCQCPGRVGTCADEKS--PCQNPCHGAAPCRVLPFG--GAQCECPFGR 1505
QY 231 SGLRCEEGVTTTSLGAGSYLALPALTNTHIEL-----LDVEFPKLAP 276
Db 1506 EGTEFCQ-----TASQDQSGSPFLA--DFNGFSHLRLGLHTFARDLGEKMALEVYFLARGP 1559
QY 277 DGVLLFSGKSGPVDFVSLAMVGGHLEFRYELGSLAVLRSAPALGRWHRVSAERLN 336
Db 1560 SGLLYNGQKTDGKDFVSLALDRERLEFRYDLKGAAVRSRPFVTLGATVSLERNG 1619
QY 337 KDGSLRVNGRPPVLRSPGKSGQLNLHTLYLGVGVEPSVPLSPATNNAHFRGCVGEVS 396
Db 1620 RKGAIRVGDGPRVLGESPVHTVNLKEPLYVGGAPDFSKLARAASVSGDFGAIQLVSL 1679
QY 397 NGKRL-----DLTYSPLGSGGICQCVDSPPCER---QPCQHGATCWPAGEYFQC 443
Db 1680 GGRQLLTFEHLVRQVDVT--SFAG-----HPCTRASHGPHCLNGASCVP----- 1720
QY 444 LCRDGFKGLDCEHENPQCLREPCLHGTTCQGTCLCLPLGFSGPRCQOQSGHGIABSDMH 503
Db 1721 -----REAY-----VCLCPGFGSPHCCKGL----- 1742
QY 504 LEGSGNDAPQOYGAYHDDGFLAPGHVFSRSLPEVPET-----ILEVRTSTASG 555
Db 1743 VEKSAG-----DVTDLAFDGRTFVEYLVNAVTESEKALQSNHNFELSRLTATQ 1790

QY 556 LLIWQGEVGEAGQKDFISLGLQDHLVFRYQLSGEARLYSEDPINDEGHRVTALRE 615
Db 1791 LVLWS---GKATERADYVALAIVDGHQLQSLNLSQPVVLRSTVPVNTNRLRVVAHEE 1846
QY 616 GRRGSLOVDGEELVSCRSPGPVNAVNAKGSVYIGAPD--VATLTGGRFPSSSITGCVKUL 673
Db 1847 QREGSLQVGNAPVTVGTSSPLGATQDLDGALWGLGLPELVGPALPKAYGTGVGCLRDV 1906
QY 674 VLHSARPGAPPQPLDLQHRQAQAGANTPCPS 705
Db 1907 VVGR-----HPLFLEDDAYTKELRCPCT 1930

RESULT 2
US-10-016-283-34
; Sequence 34, Application US/10016283
; Publication No. US20020164702A1
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/10/016,283
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US/09/077,955A
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: PCT/US96/20696
; PRIOR FILING DATE: 1996-12-13
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 34
; LENGTH: 1940
; TYPE: PRT
; ORGANISM: Rattus sp.
US-10-016-283-34

Query Match 22.5%; Score 862.5; DB 13; Length 1940;
Best Local Similarity 32.0%; Pred. No. 2.8e-61;
Matches 239; Conservative 92; Mismatches 280; Indels 135; Gaps 22;

QY 2 IKITRPSADGMILYKQKRVPGSPPTNLARQDFTSFGLVGGRPEFRFDAGSGMATIR 61
Db 1287 LALFRALETEGLLYNGNA-----RGKDFLALALDDGRVQTRDFTSGSPAVLT 1335
QY 62 HPTPLALGHFTVTLRLSLTQGLSIVGDLAPVNGTSQKFGQGLDNEELYLGVPDYG-- 119
Db 1336 SLVPEPGRWHRLELSRHRWGTLSDVGETPVLGESPGTDLNLDNLYVGGIPEEQVA 1395
QY 120 -AIPKAGLSGFGICVRELRTQGEIVFHDNLNLT--HGI\$HC--PTCRDRPCQNGQC 173
Db 1396 MYLDRTSVGVGLKGCIRMLDINNQOLELSDMQRAAVQSSGVGECGDHPCLNPCHGGALC 1455
QY 174 HDSESSYVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGL 233
Db 1456 QALEAGMFLCQCPGRVGTCADEKS--PCQNPCHGAAPCRVLSG--GAKCECPGRSGT 1513
QY 234 RCEEGVTTTSLSAG-----SYLALPALTNTHIEL-----LDVEFPKLAPDG 278
Db 1514 FCQ---TV-----LETAGSRPFLADPNFGFSYLELKLGLHTFERDLGEKMALEVFLARGPSG 1566
QY 279 VLLFSGKSGPVDFVSLAMVGGHLEFRYELGSLAVLRSAPALGRWHRVSAERLNKD 338
Db 1567 LLLYNGQKTDGKDFVSLALDRERLEFRYDLKGAAVRSRPFVTLGATVSLERNGR 1626
QY 339 GSLRVNGRPPVLRSPGKSGQ-----LNHTLLYLGVEPSVPLSPATNNAHFRGCVGE 393
Db 1627 GALQVGDGPRVLGESP--KSRKVPHTMLNKEPLYIGAPDFSKLARGAAVSSGFSGVIQL 1685
QY 394 VSVNGKRLDLYTSLGSGGICQCVDSPPCERQPCQHGATCWPAGEYFQCCLRDGFKGLD 453
Db 1686 VSLRGHQL-----LTQEHVLRADVDSFPADHPCTQA----- 1716
QY 454 CSEHENPQCLREPCLHGTGTC---QGT--RCLCLPGFSGPRCQOQSGHGIABSDMHLESGG 509

Db 1717 -----LGNPCLNGSCVREATYECCLCPGSGSLHCEKGL-----VEKSVG 1757
 Qy 510 NDAPGQYAYFHDGDFLAPFGHVSRLPEVPEP-----IELEVRTSTASGLLLWOG 561
 Db 1758 -----DLETLAPDGRITYEYLVNAVISEKALQSNFELSRLATEQGLVW-- 1803
 Qy 562 VEYGEAGQCKDFISLQLODHLVFRYQLGSGEARLVSEDPINDGEWHRVLTALREGGCSI 621
 Db 1804 --IGKAAERADYALVAIDVGHQLSLVDLGSQPVLRSTVKVNTNWLIRAHREHREGSL 1861
 Qy 622 QVDGEELVSGRSGPNNVNAKSGSVYIGAP--DVATLTGGRFFSGITGCVKNLVLSHAR 679
 Db 1862 QVNEAPVTGSSPLGATQDLDTCALWGLQKLPVGOALPKAYGTGFGVGLRDVVVGH-- 1919
 Qy 680 PGAPPPPLDLQHRAGAGANTRPCCS 705
 Db 1920 -----RQLHLEDAVTKPELRPCPT 1939

RESULT 3

US-10-108-260A-4433
 ; Sequence 4433, Application US/10108260A
 ; Publication No. US20040005560A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HELIX RESEARCH INSTITUTE
 ; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
 ; FILE REFERENCE: H1-A0106
 ; CURRENT APPLICATION NUMBER: US/10/108,260A
 ; CURRENT FILING DATE: 2002-03-27
 ; NUMBER OF SEQ ID NOS: 5458
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 4433
 ; LENGTH: 775
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-108-260A-4433

Query Match 22.0%; Score 840; DB 15; Length 775;
 Best Local Similarity 29.6%; Pred. No. 5.4e-60;
 Matches 214; Conservative 107; Mismatches 257; Indels 144; Gaps 20;
 Qy 1 EIKITFRPDSADGMLLYNQKRVPGSPTNLANRQDPDFISGLVGRPERFRFDAGSGMATI 60
 Db 176 QITLFRAEADGMLLYCENE-----HGRGDFNSLAIIRSLQFRNCGTGVAIL 226
 Qy 61 RHPTPLALGHFTVTLRLSLTQGLSVIGDLAPVNGTSQKFGQLDLNEELYGYDP-YG 119
 Db 227 VSETKIKLGGWHTVMLYRDGLNGLQLNNGTPTVTSQSQYKSKITFRTPLYLGAPSAW 286
 Qy 120 AIPKAGLSSGFTGCVELRIQGEIEIVFHDNL-----LTAHGISHCPT--CRDRPCQNG 170
 Db 287 LVRTATGTRNGFQCGVQSLAVNGRRI---DMRPWPLGKALSADVGECSSGICDEASCIHG 343
 Qy 171 GOCHDSESSYVVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
 Db 344 GTCATKADSYICLCPGLF-----362
 Qy 231 SGLRCEGVVTTPLS--SGAGSYLALPALNTNTHL-----RLDVEFKPLADGVLLFS-- 283
 Db 363 KGRHCEDAFTLTIPOFRSLRSYAATPWPLEPQHYLSFTEPEITFRPDSDGVLLSYDT 422
 Qy 284 GKSGPVEDFVSLAMVGHLFRYELGSLAVLSAEPLALGRWHRVSAERLNKDGSLRV 343
 Db 423 GSK-----DFLSINLAGHVEFRPCSGTGVLSRSEDPLTIGNWHEL RVSTAKNGILOV 477
 Qy 344 NGRVRLRSSPKSGSLMLHTLLYLGVEVPVLSPATNNAHPRGCVGEVSVNGKRLDL 403
 Db 478 DKQKIVEGMAEGFTQIKNTDIFIGFVNPVDDVKNSGVLPKPSGSGTQKIILNDRTIHW 537
 Qy 404 TVSFLLSGOGICQYDSSPCERQPOHGTATMPAGRYEQCLCRDGKGLCEHENFQCL 463
 Db 538 KHDP--TSGVNVENNAHPCVAPACAHGSGCAPRKE-GYDCCDCLPGLFEG-----582

Qy 464 REPCLHGTCQGT--RCLCLPFGSPRCQQSGHGIAESDWHLEGGSGNDAPQOYGAYFH 521
 Db 583 -----LH---CCKAIIIEAIEIPQFIG-----RSYLTYN 608
 Qy 522 DDGFLAPFGHVSRLPEVPEPTELEVRTSTASGLLLWQGEVGEAGQCKDFISLGLQDG 581
 Db 609 PDILKRVSG--SRS-----NVFMRFKTTAKDGLLLWRG--DSPMRPNDSDFISLGLRDG 657
 Qy 582 HLVPFVYOLGSGEARLVSEDPINDGEWHRVLTALREGRSGSIQVDGEBELVSGRSGPNNVAVN 641
 Db 658 ALVFSYNLGSVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDYDYGARTKSPQWRQLN 717
 Qy 642 AKGSVYIGAPDVATLTGGRFFSGITGCVKNLVLSHARPCAPPPPLDLQHRAGAGANTR 701
 Db 718 INGALVYGGMKETALHTNRTQIMRGLVGCISHFTLST-----DYHISLVEDAVDGNIN 770
 Qy 702 PC 703
 Db 771 TC 772

RESULT 4

US-09-866-050A-503
 ; Sequence 503, Application US/09866050A
 ; Publication No. US20030040471A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Watson, James D.
 ; APPLICANT: Strachan, Lorna
 ; APPLICANT: Sleeman, Matthew
 ; APPLICANT: Onrust, Rene
 ; APPLICANT: Murison, James G.
 ; APPLICANT: Kumble, Krishanand D.
 ; TITLE OF INVENTION: Compositions Isolated From Skin Cells
 ; FILE REFERENCE: 11000.1011c4U
 ; CURRENT APPLICATION NUMBER: US/09/866,050A
 ; CURRENT FILING DATE: 2001-05-24
 ; NUMBER OF SEQ ID NOS: 725
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 503
 ; LENGTH: 819
 ; TYPE: PRT
 ; ORGANISM: Rat
 US-09-866-050A-503

Query Match 21.8%; Score 832; DB 10; Length 819;
 Best Local Similarity 29.3%; Pred. No. 2.6e-59;
 Matches 210; Conservative 109; Mismatches 264; Indels 134; Gaps 17;
 Qy 1 EIKITFRPDSADGMLLYNQKRVPGSPTNLANRQDPDFISGLVGRPERFRFDAGSGMATI 60
 Db 220 QITLFRAEADGMLLYCENE-----HGRGDFNSLAIIRSLHFRNCGTGMAT 270
 Qy 61 RHPTPLALGHFTVTLRLSLTQGLSVIGDLAPVNGTSQKFGQLDLNEELYGYDP-YG 119
 Db 271 ISETKIKLGAWSVLYRDLNGLQLNNGTPTVTSQSQYKSKITFRTPLYLGAPSAW 330
 Qy 120 AIPKAGLSSGFTGCVELRIQGEIEIVFHDNL-----LTAHGISHCPT--CRDRPCQNG 170
 Db 331 LVRTATGTRNGFQCGVQSLAVNGKRI---DMRPWPLGKALNGADVGECSSGICDEASCIHG 387
 Qy 171 GOCHDSESSYVVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
 Db 388 GTCATKADSYICLCPGLFGR-----409
 Qy 231 SGLRCEGVVTTPLS--SGAGSYLALPALNTNTHL-----RLDVEFKPLADGVLLFSGKG 286
 Db 410 ---HCEDAFTLTIPOFRSLRSYAATPWPLEPQHYLSFTEPEITFRPDSDGVLLSYDT 466
 Qy 287 SGPVEDFVSLAMVGHLFRYELGSLAVLSAEPLALGRWHRVSAERLNKDGSLRVNGG 346
 Db 467 SS--KDFLSIIMAGHVEFRPCSGTGVLSRSEDPLTIGNWHLRVSTAKNGILOVDKQ 524

Db 326 RG--DSPMRNSDFISLGLDGLVSYNLSGVSIMWNGSFNDGRWHRVAVRDQSG 383
Qy 620 SIQVDEBELVSGSPGNVAVNAKSGVYIGGAPDVATLTGGRFSSGITGCVKNLVLSAR 679
Db 384 KITVDDYGARTGSPGMROLNGLYVGMKEIALHTNRQYMRGLVGCISHFTLST-- 441
Qy 680 PGAPPPPOPLDQRAQAAGANTRPC 703
Db 442 -----DYHISLVEDAVDGKNINTC 460

RESULT 7
US-09-764-853-541
; Sequence 541, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: FJ206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 541
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-853-541

Query Match 15.8%; Score 605; DB 9; Length 432;
Best Local Similarity 29.0%; Pred. No. 4.5e-41;
Matches 157; Conservative 75; Mismatches 188; Indels 122; Gaps 15;

Qy 171 GOCHDESSSYVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
Db 1 GTCTAIAKADSYICLPLGF----- 19

Qy 231 SGLRCEGVTVTPSL--SGAGSYALPALNTNHEL---RLDVEFKPLADGVLLFS--- 283
Db 20 KGRHCEDAFTLTIPQFRESLSYAATPWPLEPQHYLSFMEFEITFRPDSGDGVLLYSYDT 79

Qy 284 GKSGPVEDFVSLAMVCGHLEFRYELGSLAVLSAEPLALGRWHRVSAERLNKDGSLRV 343
Db 80 GSK-----DFLSINLAGHVEFRDCCSGTGLVRSDEPLTLGNWHELVRVSRKNGILQV 134

Qy 344 NGRPVLRSPPGKSQGLNLHTLLYLGGVEPSVPLSPATNNSAHFRGCVGEVSNKRLDL 403
Db 135 DKQKIVEGMAEGGFTQIKNTDIFIGVENVDDVKNSGVLKPFSSIQKIILNDRTHV 194

Qy 404 TYSFLSGQIGQCYDSSPCERQCHGATCMPAGEYEFQCLCRDGFKGDLCHEENPCQL 463
Db 195 KHDF--TSGVNVENAAHPCVRAPCAHGGSCRPRKE--GYDCDCPLGFEG----- 239

Qy 464 REPCLHGCTCGT--RCLCLPFGSPRCQGGSHGIAESDWHLEGGGNDAPQGYGFH 521
Db 240 ----LH---CQKAIIEAIEIPQIG-----RSYLTYN 265

Qy 522 DDGFLAPGHVFSRSLPEVETIEVRTSTAGLLMQVEVGEAGQKDFISLGLQDG 581
Db 266 PDILKRVSG---SRS-----NVFMRFKTTAKDGLLWRG--DSPMRPNSDFISLGLRDG 314

Qy 582 HLVPYQLSGEARLYSEDPINDGEWHRVTALREGRGSIQVDGEELVSGRSPGNVAVN 641
Db 315 ALVFSYNLGSGVASIMWNGSFNDGRWHRVAVRDQSGKITVDDYGARTGKSPGMROLN 374

Qy 642 AKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLSARPGAPPPQPLDQRAQAAGANTR 701
Db 375 INGLYVGMKEIALHTNRQYMRGLVGCISHFTLST-----DYHISLVEDAVDGKNIN 427

Qy 702 PC 703

702 PC 703

Db 428 TC 429

RESULT 8

US-09-764-898-179
; Sequence 179, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: FJ201
; CURRENT APPLICATION NUMBER: US/09/764,898
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 311
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 179
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-898-179

Query Match 15.8%; Score 605; DB 9; Length 432;

Best Local Similarity 29.0%; Pred. No. 4.5e-41;

Matches 157; Conservative 75; Mismatches 188; Indels 122; Gaps 15;

Qy 171 GOCHDESSSYVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
Db 1 GTCTAIAKADSYICLPLGF----- 19

Qy 231 SGLRCEGVTVTPSL--SGAGSYALPALNTNHEL---RLDVEFKPLADGVLLFS--- 283
Db 20 KGRHCEDAFTLTIPQFRESLSYAATPWPLEPQHYLSFMEFEITFRPDSGDGVLLYSYDT 79

Qy 284 GKSGPVEDFVSLAMVCGHLEFRYELGSLAVLSAEPLALGRWHRVSAERLNKDGSLRV 343
Db 80 GSK-----DFLSINLAGHVEFRDCCSGTGLVRSDEPLTLGNWHELVRVSRKNGILQV 134

Qy 344 NGRPVLRSPPGKSQGLNLHTLLYLGGVEPSVPLSPATNNSAHFRGCVGEVSNKRLDL 403
Db 135 DKQKIVEGMAEGGFTQIKNTDIFIGVENVDDVKNSGVLKPFSSIQKIILNDRTHV 194

Qy 404 TYSFLSGQIGQCYDSSPCERQCHGATCMPAGEYEFQCLCRDGFKGDLCHEENPCQL 463
Db 195 KHDF--TSGVNVENAAHPCVRAPCAHGGSCRPRKE--GYDCDCPLGFEG----- 239

Qy 464 REPCLHGCTCGT--RCLCLPFGSPRCQGGSHGIAESDWHLEGGGNDAPQGYGFH 521
Db 240 ----LH---CQKAIIEAIEIPQIG-----RSYLTYN 265

Qy 522 DDGFLAPGHVFSRSLPEVETIEVRTSTAGLLMQVEVGEAGQKDFISLGLQDG 581
Db 266 PDILKRVSG---SRS-----NVFMRFKTTAKDGLLWRG--DSPMRPNSDFISLGLRDG 314

Qy 582 HLVPYQLSGEARLYSEDPINDGEWHRVTALREGRGSIQVDGEELVSGRSPGNVAVN 641
Db 315 ALVFSYNLGSGVASIMWNGSFNDGRWHRVAVRDQSGKITVDDYGARTGKSPGMROLN 374

Qy 642 AKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLSARPGAPPPQPLDQRAQAAGANTR 701
Db 375 INGLYVGMKEIALHTNRQYMRGLVGCISHFTLST-----DYHISLVEDAVDGKNIN 427

Qy 702 PC 703
Db 428 TC 429

RESULT 9

US-09-764-881-102
; Sequence 102, Application US/09764881
; Publication No. US20030125246A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PT207
 ; CURRENT APPLICATION NUMBER: US/09/764,881
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - refer to PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 192
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 102
 ; LENGTH: 432
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-764-881-102

Query Match 15.8%; Score 605; DB 10; Length 432;
 Best Local Similarity 29.0%; Pred. No. 4.5e-41;
 Matches 157; Conservative 75; Mismatches 188; Indels 122; Gaps 15;
 QY 171 GCHDSSESSVVCVCPAGFTGSRCEHSOALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
 Db 1 GTCTAIAKADSYICLCPLGF----- 19
 QY 231 SGLRCEGVTVTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFS--- 283
 Db 20 KGRHCEADFTITIPQRESLSRYAATPWLEPQHYLSFMEFEITFRPDSGGOVLLYSYDT 79
 QY 284 GKSGPVEDFVSLAMVGGHLEFRYELGSLAVLSAEPLALGRWHRVSAERLNKOGSLRV 343
 Db 80 GSK-----DFTLSINLAGGHVEFRDCSGTGVLRSEDPDLGNWHELVRVSRKAKNGILQV 134
 QY 344 NGGRPVLRSPPKSGQLNLHTLLYLGGVPSVPLSPATNMSAHRGCGVEVSUNGKRLDL 403
 Db 135 DKQKIVEGMABEGFTQIKCNDDIFIGGVNDDVKNSGVLPKPSGSIQKILINDRTHV 194
 QY 404 TYSFLGSGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCOL 463
 Db 195 KHDF--TSGVNVENAAHPCVAPCAHGGSCRPKE-GYDCDCPLGFEG----- 239
 QY 464 REPCLHGTCQGT--RCLCLPFGSPRCQSGSGHGAESDWHLEGGSGNDAPGQYGFH 521
 Db 240 ----LH---CQKAIETAEIPIQFIG-----RSLTYDN 265
 QY 522 DDGFLAFPHGVFSRLPEVPTIELEVTSTASGLLLWQGVGEAGQKDFISLGLQDG 581
 Db 266 PDILKRVSG---SRS-----NVFMRFTTAKDGLLWRG--DSPMRPNSDFFISLGLRDG 314
 QY 582 HLVPYQLGSGEARLVSEDPINDGEWHRVTALREGRGSIQVDGEELVSGSPGNVAVN 641
 Db 315 ALVFSYNLGSGVASIMVNGSFNDGWRHVRVKAVRDQSGSKITVDYDYGARTGKSPGMWRQLN 374
 QY 642 AKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLSARPGAPPPQPLDLQHRQAQGANTR 701
 Db 375 INGALYVGMEKIALHTNRQYMRGLVGCISHFTLST-----DYHISLVEDAVDGKXIN 427
 QY 702 PC 703
 Db 428 TC 429

RESULT 10
 US-10-073-865-78
 ; Sequence 78, Application US/10073865
 ; Publication No. US20030044904A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PJ209CI
 ; CURRENT APPLICATION NUMBER: US/10/073,865
 ; CURRENT FILING DATE: 2002-02-14
 ; Prior Application removed - See file wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 154
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 78
 ; LENGTH: 432

; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-073-865-78
 Query Match 15.8%; Score 605; DB 14; Length 432;
 Best Local Similarity 29.0%; Pred. No. 4.5e-41;
 Matches 157; Conservative 75; Mismatches 188; Indels 122; Gaps 15;
 QY 171 GCHDSSESSVVCVCPAGFTGSRCEHSOALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
 Db 1 GTCTAIAKADSYICLCPLGF----- 19
 QY 231 SGLRCEGVTVTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFS--- 283
 Db 20 KGRHCEADFTITIPQRESLSRYAATPWLEPQHYLSFMEFEITFRPDSGGOVLLYSYDT 79
 QY 284 GKSGPVEDFVSLAMVGGHLEFRYELGSLAVLSAEPLALGRWHRVSAERLNKOGSLRV 343
 Db 80 GSK-----DFTLSINLAGGHVEFRDCSGTGVLRSEDPDLGNWHELVRVSRKAKNGILQV 134
 QY 344 NGGRPVLRSPPKSGQLNLHTLLYLGGVPSVPLSPATNMSAHRGCGVEVSUNGKRLDL 403
 Db 135 DKQKIVEGMABEGFTQIKCNDDIFIGGVNDDVKNSGVLPKPSGSIQKILINDRTHV 194
 QY 404 TYSFLGSGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCOL 463
 Db 195 KHDF--TSGVNVENAAHPCVAPCAHGGSCRPKE-GYDCDCPLGFEG----- 239
 QY 464 REPCLHGTCQGT--RCLCLPFGSPRCQSGSGHGAESDWHLEGGSGNDAPGQYGFH 521
 Db 240 ----LH---CQKAIETAEIPIQFIG-----RSLTYDN 265
 QY 522 DDGFLAFPHGVFSRLPEVPTIELEVTSTASGLLLWQGVGEAGQKDFISLGLQDG 581
 Db 266 PDILKRVSG---SRS-----NVFMRFTTAKDGLLWRG--DSPMRPNSDFFISLGLRDG 314
 QY 582 HLVPYQLGSGEARLVSEDPINDGEWHRVTALREGRGSIQVDGEELVSGSPGNVAVN 641
 Db 315 ALVFSYNLGSGVASIMVNGSFNDGWRHVRVKAVRDQSGSKITVDYDYGARTGKSPGMWRQLN 374
 QY 642 AKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLSARPGAPPPQPLDLQHRQAQGANTR 701
 Db 375 INGALYVGMEKIALHTNRQYMRGLVGCISHFTLST-----DYHISLVEDAVDGKXIN 427
 QY 702 PC 703
 Db 428 TC 429

RESULT 11
 US-10-242-747-102
 ; Sequence 102, Application US/10242747
 ; Publication No. US20040005577A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PT207CI
 ; CURRENT APPLICATION NUMBER: US/10/242,747
 ; CURRENT FILING DATE: 2002-09-13
 ; Prior Application removed - See file wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 154
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 102
 ; LENGTH: 432

; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 192
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 102
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-242-747-102

Query Match 15.8%; Score 605; DB 15; Length 432;
Best Local Similarity 29.0%; Pred. No. 4.5e-41;
Matches 157; Conservative 75; Mismatches 188; Indels 122; Gaps 15;
QY 171 GOCHDSESSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGVTCTCHLGR 230
Db 1 GTCTAIKADSYICLPLGF----- 19
QY 231 SGLRCEGVTTTPSL-SGAGSYALPALPNTNTHHEL---RLDVEFKPLAPDGVLLFS--- 283
Db 20 KGRHCEDAFTLTPQFRESLSRYAATFWLEPQHLYSFMFEFTTFRPDSGDGVLLSYDT 79
QY 284 GKSGPVEDFVSLAMVGGHLEFVYELGSLAVLSAEPLALGRWHRVSAERLNKDGSLRV 343
Db 80 GSK-----DPLSNLAGHVEFPDGGTGVLRSEDPTLGNWHELVRSTAKGILQV 134
QY 344 NGRPVLRSPPGKSGQGLNHLTLVYLGVEFSPVLSPATNWSAHFRGCVGEVSVNGKRLD 403
Db 135 DKOKIVEGMAEGGFTQIKCNTDIFIGVNPYDDVKNKSGVLKPFSGSIQKIILNDRTHIV 194
QY 404 TYSFLSGQIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGLCHEENPQOL 463
Db 195 KHDF--TSGVNVENAAHPCVAPCAHGGSCRPKE-GYDCDCPLGPEG----- 239
QY 464 REPCILHGTCQGT--RCLCLPFGSGPCQCGSGHGAESDWHLESGGNDAPGQYCAVPH 521
Db 240 ----LH---CQKAIIEAIEIPQIG-----RSYLYTDN 265
QY 522 DCGFLAPPGHVFRSRLPEVPETIELEVRTSTASGLLWQGVGEAGQCKDFISLQDQG 581
Db 266 PDILKRVSG--SR-----NVFMFKTKAKDGLLNRG--DSPMRNSDFISLGRDG 314
QY 582 HLVFRYQLGSGEARLYSEDPINDGEHVRTALREGRGSIQVDGELVSGRSPGNVAVN 641
Db 315 ALVFSYNLGSGVASIWNVSFNDGRHVRKAVRDGSGKLTVDYDYGARTGKSPGMRQLN 374
QY 642 AKGSVYIGGAPDVATITGGRFSSGITGCVKNLVLSHARPAPPPQPLDLOHRAQAQANTR 701
Db 375 INGALYVGGMKEITALHTRNQYMRGLVGCISHTLST-----DVHISLVEDAVDGNKIN 427
QY 702 PC 703
Db 428 TC 429

RESULT 12
US-10-016-283-25
; Sequence 25, Application US/10016283
; Publication No. US20020164702A1
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/10/016,283
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US/09/077,955A
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: PCT/US96/20696

; PRIOR FILING DATE: 1996-12-13
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 25
; LENGTH: 456
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-016-283-25

Query Match 15.2%; Score 579.5; DB 13; Length 456;
Best Local Similarity 32.7%; Pred. No. 6e-39;
Matches 168; Conservative 57; Mismatches 156; Indels 133; Gaps 16;
QY 242 TTPSLSGAGSYALPALPNTNTHHEL---LDVEFKPLAPDGVLLFSGSKS 287
Db 25 TASQDGSGLPLA-DNFGFSLHLEURLHTFARDIGERKVALEVEFLARGPSGLLYNGOKT 83
QY 288 GPVEDFVSLAMVGGHLEFVYELGSLAVLSAEPLALGRWHRVSAERLNKDGSLRVNGGR 347
Db 84 DGKGFVSLALRDRRLEFRYDLGKGAIVRSREPTVLTGATRVSLERNRKGALRVGDGP 143
QY 348 PVLRSPPGKSGQ-----LNHLTLVYLGVEFSPVLSPATNWSAHFRGCVGEVSVNGKRL- 401
Db 144 RVLGESP-KSRKVPHTVNLKEPLYVGGAPDFSKLAPAAAIVSSGFDGAIQLVSLGSRQLL 202
QY 402 -----DLTYSFLSGQIGQCYDSSPCER--QPCQHGATCMPAGEYEFQCLCRDGF 449
Db 203 TPEHVLRAQDVT-SFAG-----HPCTRASGHECLNGASCVP----- 237
QY 450 KGDLCHEHENPCQUREPCLHGTCQGTCTCLCPFGSPRCQCGSGHGAESDWHLESGGG 509
Db 238 -----REAY-----VCLCPGFGSGPHCEKGL-----VETKSAG 265
QY 510 NDAPGQYGAFFHDDGFLAHPGHVFRSRLPEVPET-----LELEVRTSTA 553
Db 266 -----DVDTLAFDGRTFEVLNAVYSELANEIPVEKALQSNHFEISLRTAT 313
QY 554 SGLLWQGVGEAGQCKDFISLQDGHVFRYQLGSGEARLYSEDPINDGEHVRTAL 613
Db 314 QGLVWS---GKATERADYVALAIVDHLQLSLYNLGSQPWLRSTVFNTRNRLRVVAH 369
QY 614 REGRGSIQVDGELVSGRSPGNVAVNAGSVYIGGAPD--VATLTGGRFSSGITGCVK 671
Db 370 REQREGSLQVNEAPVTSGLSGATQDLDGALWGLGPELVPGLPKAYGTGFGVGLCLR 429
QY 672 NLVLHSARPAPPPQPLDLOHRAQAQANTRPCPS 705
Db 430 DVVYGR-----HPLHLLDVAVKPELRCPPT 455

RESULT 13
US-10-016-283-26
; Sequence 26, Application US/10016283
; Publication No. US20020164702A1
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/10/016,283
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US/09/077,955A
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: PCT/US96/20696
; PRIOR FILING DATE: 1996-12-13
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 26
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-016-283-26

Query Match 15.1%; Score 576; DB 13; Length 440;

Db 251 VLMS---GKATERADYVALAIVDGHLLQLSYNLSQSPVLRSTVPVNTNRWLRVVAHREQ 306
Qy 617 RRSIQVDGBELVSGRSPGNVAVNAXGSVYICGAPD--VATLTGGRFSSGITGCVKNLV 674
Db 307 REGSLQVGNAPVTGSSPLGATOLDTDGALWLGCLPELPVGPALPKAYGTGFVGCCLRDVV 366
Qy 675 LHSARPGAPPPQPLDLQHRAQAGANTPCPS 705
Db 367 VGR-----HPLHLLLEDAVTKPELRPCPT 389

Search completed: March 9, 2004, 17:25:14
Job time : 54.7431 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:16:23 ; Search time 27.6355 Seconds ~
(without alignments)
1317.011 Million cell updates/sec

Title: US-10-006-011A-3

Perfect score: 3825

Sequence: 1 EIKITRPDSADGMLYNGQ.....QPLDLQHRQAQAGANTRPCPS 705

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/prodata/2/iaa/5A COMB.pcp.*
2: /cgn2_6/prodata/2/iaa/5B COMB.pcp.*
3: /cgn2_6/prodata/2/iaa/6A COMB.pcp.*
4: /cgn2_6/prodata/2/iaa/6B COMB.pcp.*
5: /cgn2_6/prodata/2/iaa/PCUTUS COMB.pcp.*
6: /cgn2_6/prodata/2/iaa/backfile1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	862.5	22.5	1940	2	US-08-644-271-30
2	862.5	22.5	1940	4	US-09-077-955-34
3	640.5	16.7	492	2	US-08-644-271-32
4	640.5	16.7	492	4	US-09-077-955-36
5	579.5	15.2	456	4	US-09-077-955-25
6	576	15.1	440	4	US-09-077-955-26
7	528	13.8	390	4	US-09-077-955-27
8	397.5	10.4	771	3	US-09-188-930-183
9	397.5	10.4	1525	3	US-09-191-647-2
10	397.5	10.4	1525	3	US-09-540-245A-2
11	397.5	10.4	1525	3	US-08-540-153-2
12	394.5	10.3	1523	4	US-09-312-283C-396
13	394	10.3	716	4	US-09-312-283C-183
14	391.5	10.2	1523	4	US-09-182-024A-2
15	382	10.0	338	4	US-09-077-955-28
16	361	9.4	3075	2	US-08-460-309-5
17	361	9.4	3075	2	US-08-125-077-5
18	359	9.4	1130	6	5444158-2
19	357	9.3	1130	2	US-08-460-309-2
20	357	9.3	1130	2	US-08-125-077-2
21	357	9.3	3088	4	US-09-562-702A-8
22	357	9.3	3089	4	US-09-562-702A-4
23	357	9.3	3110	4	US-09-562-702A-2
24	357	9.3	3110	4	US-08-562-702A-6
25	357	9.3	3110	4	US-09-561-709B-7
26	357	9.3	3111	2	US-08-460-309-4
27	357	9.3	3111	2	US-08-125-077-4

28 349.5 9.1 735 3 US-08-191-647-9 Sequence 9, Appli
29 349.5 9.1 735 3 US-09-540-245A-9 Sequence 9, Appli
30 349.5 9.1 735 3 US-09-540-153-9 Sequence 9, Appli
31 347 9.1 3084 4 US-09-562-702A-12 Sequence 12, Appli
32 347 9.1 3106 4 US-09-562-702A-10 Sequence 10, Appli
33 340 8.9 294 4 US-09-077-955-29 Sequence 29, Appli
34 333 8.7 1480 3 US-09-191-647-7 Sequence 7, Appli
35 333 8.7 1480 3 US-09-540-245A-7 Sequence 7, Appli
36 333 8.7 1480 3 US-09-540-153-7 Sequence 7, Appli
37 333 8.7 1480 5 PCT-US91-09055-2 Sequence 2, Appli
38 328.5 8.6 2471 1 US-08-185-432-16 Sequence 16, Appli
39 328.5 8.6 2471 1 US-08-083-590A-19 Sequence 19, Appli
40 328.5 8.6 2471 3 US-08-532-384-19 Sequence 19, Appli
41 328.5 8.6 2471 4 US-08-899-232-1 Sequence 1, Appli
42 328 8.6 1480 4 US-08-182-024A-5 Sequence 5, Appli
43 326 8.5 256 4 US-09-077-955-30 Sequence 30, Appli
44 322.5 8.4 3571 4 US-09-911-842A-2 Sequence 2, Appli
45 316 8.3 2321 4 US-09-230-652-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-644-271-30
; Sequence 30, Application US/08644271
; Patent No. 5814478
; GENERAL INFORMATION:
; APPLICANT: Valenzuela, et al.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/644,271
; FILING DATE: 10-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: USSN 60/008,657
; FILING DATE: 15-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Cobert, Robert J
; REGISTRATION NUMBER: 36,108
; REFERENCE/DOCKET NUMBER: REG 195A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 914-345-7400
; TELEFAX: 914-345-7721
; TELEX:
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1940 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Rat Agrin
; LOCATION: 1..1940
; OTHER INFORMATION:
US-08-644-271-30

Query Match 22.5%; Score 862.5; DB 2; Length 1940;

Best Local Similarity 32.0%; Pred. No. 7.4e-64;
Matches 239; Conservative 92; Mismatches 280; Indels 135; Gaps 22;

QY 2 IKITRPDSADGMLLYNGKRVPGSPNTLANRQPDFISGLVGRPERFRDAGSGMATIR 61
DB 1287 LALFRALETEGGLLYNGNA-----RGKDFLALALLDGRVQFRFTDSSGPAVLT 1335
QY 62 HPTPLALGHFTVTLRLSLTOSGLIVGDLAPVNGTSQKFGQLDNEELYLGYPDYG-- 119
DB 1336 SLVPVEPGRWHLRSLRHWQGLTSLVDGETPVVGSPPSGTDGLNLDNLNLYVGGIPEBQVA 1395
QY 120 -AIPKAGLSSGFIGVRELRIOGEEIVFHDNLNTA---HGISHC--PTCRDRPCONGGQC 173
DB 1396 MYLRTSVGVKGCIRMLDINNQLESLDWQRAAVQSSGVGECGDHPCLPNCHGALC 1455
QY 174 HDSSESSVVCVCPAGTSGRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGL 233
DB 1456 QALEAGMFLCQCPGRFGPTCADEKS-PCQPNPCHGAAPCRVLSG-CAKCECPGRSGT 1513
QY 234 RCEEGVTTTPSLSGAG-----SYLALPALNTNTHHEL-----RLDVEFKPLADPG 278
DB 1514 FCQ---TV---LETAGSRPFLADFNFGFSYLELKGHTFERDLGKMALEWVFLARPSG 1566
QY 279 VLLFSGGSGPVEDFVSLAMVGGHLEFRYELGSLGLAVLSAEPLALGRWHRVSAERLNDK 338
DB 1567 LLLYNGQKTDGKDFVSLALNHRHLEFCYDLGKGAAVIRSKPEIALTGTWVRVFLERNGRK 1626
QY 339 GSLRVNGGRPVLRSPGKSGQ-----LNHTLLYLGVEPSVPLSPATNWSAHPGCVGE 393
DB 1627 GALQVGDGPRVIGESP-KSRKVPHTMLNKLKPLVIGGAPDFSKLARGAAVSSGFSGVQL 1685
QY 394 VSVNGKRDLTYSFILGSGICQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGLD 453
DB 1686 VSLRGHQL-----LTQEHVLRVADVSPFADHPCTQA-----1716
QY 454 CEHENPCLRPCLHGGTC---QGT-RLCLPFGSPGRCQSGHGIASDWHLESGG 509
DB 1717 -----LGNPCLNGSGCVPREATYECCLPGFSGHLHCKGL-----VEKSVG 1757
QY 510 NDAPQGYGAYFHDDGFLAPPGHVSRSLEPEVET-----IELEVRTSTASGLLLWQG 561
DB 1758 -----DLFTLAFDGRTYEYLNATIESEKALQSNHFELSRLTEATQGLVLM-- 1803
QY 562 VEVGAGQKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGEWHRVTLRGRGSI 621
DB 1804 --IGKAAERADYMALAIVDGHQLSLSYDLGSPVVLSTVKVNTNRWLRIRAHREHREGSL 1861
QY 622 QVDGEELYSRSPGNVAVNAKGSVYIGGAP--DVATITGGRFSSGITGCVKNLVLHSAR 679
DB 1862 QVNEAPVTGSSPLGATQDITDGLWLGGLQKLPVGOALPKAYGTGFCGLRDVVVGH-- 1919
QY 680 PGAPPPQPLDLQHRAQAAGANTRPCPS 705
DB 1920 -----RQLHLEDAVTKPELRPCPT 1939

RESULT 2
US-09-077-955-34
; Sequence 34, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; EARLIER FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36

SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
; LENGTH: 1940
; TYPE: PRT
; ORGANISM: Rattus sp.
US-09-077-955-34

Query Match 22.5%; Score 862.5; DB 4; Length 1940;
Best Local Similarity 32.0%; Pred. No. 7.4e-64;
Matches 239; Conservative 92; Mismatches 280; Indels 135; Gaps 22;

QY 2 IKITRPDSADGMLLYNGKRVPGSPNTLANRQPDFISGLVGRPERFRDAGSGMATIR 61
DB 1287 LALFRALETEGGLLYNGNA-----RGKDFLALALLDGRVQFRFTDSSGPAVLT 1335
QY 62 HPTPLALGHFTVTLRLSLTOSGLIVGDLAPVNGTSQKFGQLDNEELYLGYPDYG-- 119
DB 1336 SLVPVEPGRWHLRSLRHWQGLTSLVDGETPVVGSPPSGTDGLNLDNLNLYVGGIPEBQVA 1395
QY 120 -AIPKAGLSSGFIGVRELRIOGEEIVFHDNLNTA---HGISHC--PTCRDRPCONGGQC 173
DB 1396 MYLRTSVGVKGCIRMLDINNQLESLDWQRAAVQSSGVGECGDHPCLPNCHGALC 1455
QY 174 HDSSESSVVCVCPAGTSGRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGL 233
DB 1456 QALEAGMFLCQCPGRFGPTCADEKS-PCQPNPCHGAAPCRVLSG-CAKCECPGRSGT 1513
QY 234 RCEEGVTTTPSLSGAG-----SYLALPALNTNTHHEL-----RLDVEFKPLADPG 278
DB 1514 FCQ---TV---LETAGSRPFLADFNFGFSYLELKGHTFERDLGKMALEWVFLARPSG 1566
QY 279 VLLFSGGSGPVEDFVSLAMVGGHLEFRYELGSLGLAVLSAEPLALGRWHRVSAERLNDK 338
DB 1567 LLLYNGQKTDGKDFVSLALNHRHLEFCYDLGKGAAVIRSKPEIALTGTWVRVFLERNGRK 1626
QY 339 GSLRVNGGRPVLRSPGKSGQ-----LNHTLLYLGVEPSVPLSPATNWSAHPGCVGE 393
DB 1627 GALQVGDGPRVIGESP-KSRKVPHTMLNKLKPLVIGGAPDFSKLARGAAVSSGFSGVQL 1685
QY 394 VSVNGKRDLTYSFILGSGICQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGLD 453
DB 1686 VSLRGHQL-----LTQEHVLRVADVSPFADHPCTQA-----1716
QY 454 CEHENPCLRPCLHGGTC---QGT-RLCLPFGSPGRCQSGHGIASDWHLESGG 509
DB 1717 -----LGNPCLNGSGCVPREATYECCLPGFSGHLHCKGL-----VEKSVG 1757
QY 510 NDAPQGYGAYFHDDGFLAPPGHVSRSLEPEVET-----IELEVRTSTASGLLLWQG 561
DB 1758 -----DLFTLAFDGRTYEYLNATIESEKALQSNHFELSRLTEATQGLVLM-- 1803
QY 562 VEVGAGQKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGEWHRVTLRGRGSI 621
DB 1804 --IGKAAERADYMALAIVDGHQLSLSYDLGSPVVLSTVKVNTNRWLRIRAHREHREGSL 1861
QY 622 QVDGEELYSRSPGNVAVNAKGSVYIGGAP--DVATITGGRFSSGITGCVKNLVLHSAR 679
DB 1862 QVNEAPVTGSSPLGATQDITDGLWLGGLQKLPVGOALPKAYGTGFCGLRDVVVGH-- 1919
QY 680 PGAPPPQPLDLQHRAQAAGANTRPCPS 705
DB 1920 -----RQLHLEDAVTKPELRPCPT 1939

RESULT 3
US-08-644-271-32
; Sequence 32, Application US/08644271
; Patent No. 5814478
; GENERAL INFORMATION:
; APPLICANT: Valenzuela, et al.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
; TITLE OF INVENTION: AND LIGANDS
; NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Regeneron Pharmaceuticals, Inc.
 STREET: 777 Old Saw Mill Road
 CITY: Tarrytown
 STATE: NY
 COUNTRY: USA
 ZIP: 10591
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/644,271
 FILING DATE: 10-MAY-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: USSN 60/008,657
 FILING DATE: 15-DEC-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Covert, Robert J.
 REGISTRATION NUMBER: 36,108
 REFERENCE/DOCKET NUMBER: REG 195A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 914-345-7400
 TELEFAX: 914-345-7721
 TELEX:
 INFORMATION FOR SEQ ID NO: 32:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 492 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 FEATURE:
 NAME/KEY: Human Agrin
 LOCATION: 1..492
 OTHER INFORMATION:
 US-08-644-271-32

Query Match 16.7%; Score 640.5; DB 2; Length 492;
 Best Local Similarity 33.0%; Pred. No. 6.6e-46;
 Matches 183; Conservative 59; Mismatches 173; Indels 139; Gaps 18;
 US-08-644-271-32
 QY 202 CHPEACGPDATCVRPDGRGYTCRCHLGRSGRLCEBEGVTVTTPSLSGAGSYLALPALNT 261
 DB 27 CQNPCHGAAPCRVLPPEG-GAQCCEPLGREGTFCQ-----TASGDGSGPFLA-DFNGFS 79
 QY 262 HHLELR-----LDVEFKPLADPGVLLFSGGKSGPVDFVSLAMVGGHLEFRY 307
 DB 80 HLELRGLHTFARDLGEKMALEVVLARGPSGLLYNGQKTDGKDFVSLALDRLEFRY 139
 QY 308 ELGSLAVLRSAAPLALGRWHRVSAERLNKDGSLRVNGRPRVLRSSPKGSGQ-----LNL 362
 DB 140 DLGKGAIVRSRBPVTLGAWTRVSLERNRKGALRVGDGPRVLGESP-KSRKVPHTVNL 198
 QY 363 HTLLYLGGVEPSVPLSPATNMSAHFRGCGVEVSNKRL-----DLTYSFLGSGQ 412
 DB 199 KEPLYVGAPDFSKLARAASVSGFDGAIQLVSLGGRQLLTPEHVLQVDVT-SPAG--- 254
 QY 413 IGQCYDSSPCR---OPCOHGATCMPAGEYEFQCLCRDGFKDCHEHENPCQLREPC 469
 DB 255 -----HPCTRASGHPCLNGASCVP-----REAY- 278
 QY 470 GGTCTQTRCLCLPFGSPRCQSGHGIABSDWHLESGGNDAPQGYGAYFDDGFLAPP 529
 DB 279 -----VCLCPGFGSGPHCEKGL-----VEKSAG-----DVTDLAFD 309
 QY 530 GHVFSRSLPEVPT-----IELEVRTSTASGLLWQGVGEAGQKDF 573
 DB 310 GRTFEYVNLNATESELANEIPVEKALQSNHFELSLRTEATQGLVLS-----GKATERADY 365

QY 574 ISLGLQDGHVFRYQLGSGEARLVSEDPINDGEWHRTALREGRGSIQVDGEELVSGRS 633
 DB 366 VALAIVDGHQLQLSYNLGSPVVLASTVPTVNTNRLRVVAHREQREGSLQVNEAPVTGSS 425
 QY 634 PGPVAVNAKGSVYIGAPD--VATLTGGRFPSSGITGVKXNLVLSARPAGPPQPDLQ 691
 DB 426 PLGATQDITDGLWLGUPPLVPGPALPKAYGTGVGLRDRVVVGR-----HPLHL 477
 QY 692 HRAQAGANTRPCPS 705
 DB 478 EDAVTKPELRPCPT 491
 RESULT 4
 US-09-077-955-36
 ; Sequence 36, Application US/09077955A
 ; Patent No. 6413740
 ; GENERAL INFORMATION:
 ; APPLICANT: Valenzuela et al., David M.
 ; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
 ; FILE REFERENCE: REG195-B-PCT-US
 ; CURRENT APPLICATION NUMBER: US/09/077,955A
 ; CURRENT FILING DATE: 1998-09-10
 ; EARLIER APPLICATION NUMBER: PCT/US96/20696
 ; EARLIER FILING DATE: 1996-12-13
 ; EARLIER APPLICATION NUMBER: 08/644,271
 ; EARLIER FILING DATE: 1996-05-10
 ; EARLIER APPLICATION NUMBER: 60/008,657
 ; EARLIER FILING DATE: 1995-12-15
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: Patent in Ver. 2.0
 ; SEQ ID NO 36
 ; LENGTH: 492
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-077-955-36

Query Match 16.7%; Score 640.5; DB 4; Length 492;
 Best Local Similarity 33.0%; Pred. No. 6.6e-46;
 Matches 183; Conservative 59; Mismatches 173; Indels 139; Gaps 18;
 US-09-077-955-36
 QY 202 CHPEACGPDATCVRPDGRGYTCRCHLGRSGRLCEBEGVTVTTPSLSGAGSYLALPALNT 261
 DB 27 CQNPCHGAAPCRVLPPEG-GAQCCEPLGREGTFCQ-----TASGDGSGPFLA-DFNGFS 79
 QY 262 HHLELR-----LDVEFKPLADPGVLLFSGGKSGPVDFVSLAMVGGHLEFRY 307
 DB 80 HLELRGLHTFARDLGEKMALEVVLARGPSGLLYNGQKTDGKDFVSLALDRLEFRY 139
 QY 308 ELGSLAVLRSAAPLALGRWHRVSAERLNKDGSLRVNGRPRVLRSSPKGSGQ-----LNL 362
 DB 140 DLGKGAIVRSRBPVTLGAWTRVSLERNRKGALRVGDGPRVLGESP-KSRKVPHTVNL 198
 QY 363 HTLLYLGGVEPSVPLSPATNMSAHFRGCGVEVSNKRL-----DLTYSFLGSGQ 412
 DB 199 KEPLYVGAPDFSKLARAASVSGFDGAIQLVSLGGRQLLTPEHVLQVDVT-SPAG--- 254
 QY 413 IGQCYDSSPCR---OPCOHGATCMPAGEYEFQCLCRDGFKDCHEHENPCQLREPC 469
 DB 255 -----HPCTRASGHPCLNGASCVP-----REAY- 278
 QY 470 GGTCTQTRCLCLPFGSPRCQSGHGIABSDWHLESGGNDAPQGYGAYFDDGFLAPP 529
 DB 279 -----VCLCPGFGSGPHCEKGL-----VEKSAG-----DVTDLAFD 309
 QY 530 GHVFSRSLPEVPT-----IELEVRTSTASGLLWQGVGEAGQKDF 573
 DB 310 GRTFEYVNLNATESELANEIPVEKALQSNHFELSLRTEATQGLVLS-----GKATERADY 365
 QY 574 ISLGLQDGHVFRYQLGSGEARLVSEDPINDGEWHRTALREGRGSIQVDGEELVSGRS 633
 DB 366 VALAIVDGHQLQLSYNLGSPVVLASTVPTVNTNRLRVVAHREQREGSLQVNEAPVTGSS 425


```
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 27
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-077-955-27

Query Match      13.8%; Score 528; DB 4; Length 390;
Best Local Similarity 33.0%; Pred. No. 1.6e-36;
Matches 149; Conservative 50; Mismatches 134; Indels 118; Gaps 14;

QY 291 EDVSLAMVGGHLEFRYELGSLAVLSAEPLALGEWHVSAERLNKDGSLRVNGRPVL 350
DB 21 DKVSLALDRRLERFRIDYLGKGAIVIRSEPTVLGATRVSLERNRKGALRVGDPVRL 80
QY 351 RSPGKSGQ-----LMLHTLYLGGVEPSVPLSPATNMSAHFRGCGVSVNGKEL----- 401
DB 81 GESP-KSRKVPHTVNLKEPLVGGAPDFSLARAAVSSGFDGAIQLVSLGGRQLTPE 139
QY 402 -----DLTYSLGSGOICQVDSPCR-----OPQHGNATCMPAGEYEFQCLCRDGFKGD 452
DB 140 HVLQVDVT-SFAG-----HPCTRASGHFCLNGASCVP----- 171
QY 453 LCEHEENPQLREPCPLHGTCQCTRCCLPLGFSGRPCQGGSHGIAESDWHLEGGGND 512
DB 172 -----REAY-----VCLCPGSGPHCEKGL-----VEKSAG----- 199
QY 513 PQQYGAFFDDGLAPPGHVSRSLEVPET-----TELEVRTSTASGL 556
DB 200 -----DVTFLAFDGRTFVEYLNVAETSELANEIPVKALQSNHFLSLRTEATQGL 250
QY 557 LLWQGVGEAGQGDFTSLQDQHLVFRYQLGSGEARLVSEDPINDGEWHRVTALEQ 616
DB 251 VLWS-----GKATERADYVALAIVDGHQLSLYNLGSQPVVLASTVPTVNTNRMLRVVAHQ 306
QY 617 RGSTQVDGEELVSRGSPGNVAVNAKGSVYTGAPD--VATLTGCRFSSGITGVKNLV 674
DB 307 REGSLQVGNAPVTGSSPLGATQDLDGALWLGGLFELPVGVPALPKAYGTGFGVGLRDVV 366
QY 675 LHSARPGAPPPPOPLDLOHRAQAGANTRCPS 705
DB 367 VGR-----HPLHLEDAVTKBELRCPPT 389

RESULT 8
US-09-188-930-183
; Sequence 183, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 183
; LENGTH: 771
```

```
; TYPE: PRT
; ORGANISM: Rat
; FEATURES:
; NAME/KEY: UNSURE
; LOCATION: (717)...(717)
US-09-188-930-183

Query Match      10.4%; Score 397.5; DB 3; Length 771;
Best Local Similarity 27.6%; Pred. No. 5e-25;
Matches 145; Conservative 64; Mismatches 185; Indels 131; Gaps 28;

QY 158 HC-----PTCRDPCQNGGQCHDSSESSYVCVCPAGFTGSCHEHSQAL-----HCHPEAC 207
DB 315 HCDIDFDCCQNKCKNGAHCTDA-VNGYTCVCEGYSGLFCEFPSPMVFURTSPCNFDC 373
QY 208 GPDATC---VNRPDGRGYTCRCHLGRSGLRCEBEGVTVTPSLSGAGSYLALPAL-----T 259
DB 374 QNGAQCIIRVNEP-----IQCLPGVLGKCEKLVSVSI--LVNKESYLQIPSAKVRPQT 426
QY 260 NTHHELRLDVEFFKPLADPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSLA-VLRS 318
DB 427 NITQIATDED-----SGILLYKGDK-----DHIAVESIEG-IRASYDTGSHPSAISY 474
QY 319 AEPLALGRHVRSAERLNKDGSLRVNGRPLRSSPKSQCLNLTLLYLGVE-----P 373
DB 475 VETINDGNFHVVELLTLDSSLSLVDDGSPKIIITNLSKQSTLNFDSPLYVGGMPGNVA 534
QY 374 SVPLSPATNMSAHFRGCGVSVNGKRLDLYTSFLSGOIGQCYDSSPCRQPCQHGATC 433
DB 535 SLRQAPQGN-GTSFHGCIIRNLYINSLQDFRKVPMTGILPGC---EPCHKKYCAHG-TC 589
QY 434 MPAGEVEFQCLCRDGFKGDICEHEN-PCOLRBPCLHGGTCQ-----GTRCLCLPGFSGP 487
DB 590 QPSSQSGFTCECEBGMWGLCDQRTNDPC-LGNKCVH-GTCLPINAFSYCKLEHGGV 647
QY 488 RCOQSGHGIATSDWHLEGGGNDAPQYGCAYPHDDGFLAPPG-HVFSRLSLPVPETIEL 546
DB 648 LCDE-----EEDLFNPLPGDQVQAR----- 667
QY 547 EVRTSTASGLLW--QGVVEVGEAGQKDFISLQDQHLVFRYQLGSGEARLVSEDPIND 604
DB 668 EVQALWARAALLWQWQHHRGQLTQ-----RISCRGERIDYQ--SSVRCLL-----ND 716
QY 605 GEHVRVTA-----LREGR-----GSIQVDGEELY 629
DB 717 KEVSRLECRGCAGCGCCGCLRSKRRKYSFECTDGSFVDEVEKV 761

RESULT 9
US-09-191-647-2
; Sequence 2, Application US/09191647
; Patent No. 6046015
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: 398-031-3
; CURRENT APPLICATION NUMBER: US/09/191,647
; CURRENT FILING DATE: 1998-11-13
; EARLIER APPLICATION NUMBER: 60/065,544
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: 60/081,057
; EARLIER FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: human
US-09-191-647-2
```

Query Match 10.4%; Score 397.5; DB 3; Length 1525;
Best Local Similarity 29.7%; Pred. No. 1.3e-24;
Matches 108; Conservative 58; Mismatches 141; Indels 57; Gaps 18;

Qy 158 HC-----PTCRDRPCQNGQCHDSESSSYVCVPAGFTGSRCEHSQAL-----HCHPEAC 207
Db 1068 HCDIDFDDCQDNKCKNGAHCCTDA-VNGYTCICPEGYSGLCFCEFPWVLPRTSPCDNFD 1126

Qy 208 GPDATC---VNRPDGRGYTCRHLGRSGLRCEGVTTPSLSGAGSYLALPAL-----T 259
Db 1127 QNGAQCIVRINEP-----ICQCLPGYQGEKELVSV---NFINKESYLQIPSAKVRPQT 1178

Qy 260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEFVSLAMVGGHLEFRYELGSLA-VLRS 318
Db 1179 NITLIQIATDED-----SGILLYKGDK-----DHIAVELYRGRVRSYDTGSHPASAIYS 1227

Qy 319 AEPLALGRWHRVSAERLNKQGLSVNGRVPVLRSSPKSQGLNHLTLILYLGVE-----P 373
Db 1228 VETINDGNFHIIVELLALDQSLSVSDGNGPKIITNLSKQSTLNFDSPLYYGGMPGKSNVA 1287

Qy 374 SVPLSPATNMSAHRGCVGVSVNGKRLDLTYSFLSGQIGQCQYDSSPCERQPCQHGATC 433
Db 1288 SLRQAPQN-GTSPHGICIRNLYINSELQDFQKVPMTGILPGC---EPCHKVCAGH-TC 1342

Qy 434 MPAGEYFQCLCRDGFKGDLCEHEEN-PCQLRBPCLHGGTCQ-----GTRCICLPFGSGP 487
Db 1343 QPSSQAGFTCECQEGWGMPLCDORTNDPC-LGNKCVH-GTCLPINAFSYSCKLEHGGV 1400

Qy 488 RCOQ 491
Db 1401 LCDE 1404

RESULT 10
US-09-540-245A-2
; Sequence 2, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; CURRENT APPLICATION NUMBER: US/09/540,245A
; CURRENT FILING DATE: 2000-03-31
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: human
US-09-540-245A-2

Query Match 10.4%; Score 397.5; DB 3; Length 1525;
Best Local Similarity 29.7%; Pred. No. 1.3e-24;
Matches 108; Conservative 58; Mismatches 141; Indels 57; Gaps 18;

Qy 158 HC-----PTCRDRPCQNGQCHDSESSSYVCVPAGFTGSRCEHSQAL-----HCHPEAC 207
Db 1068 HCDIDFDDCQDNKCKNGAHCCTDA-VNGYTCICPEGYSGLCFCEFPWVLPRTSPCDNFD 1126

Qy 208 GPDATC---VNRPDGRGYTCRHLGRSGLRCEGVTTPSLSGAGSYLALPAL-----T 259
Db 1127 QNGAQCIVRINEP-----ICQCLPGYQGEKELVSV---NFINKESYLQIPSAKVRPQT 1178

Qy 260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEFVSLAMVGGHLEFRYELGSLA-VLRS 318
Db 1179 NITLIQIATDED-----SGILLYKGDK-----DHIAVELYRGRVRSYDTGSHPASAIYS 1227

Qy 319 AEPLALGRWHRVSAERLNKQGLSVNGRVPVLRSSPKSQGLNHLTLILYLGVE-----P 373
Db 1228 VETINDGNFHIIVELLALDQSLSVSDGNGPKIITNLSKQSTLNFDSPLYYGGMPGKSNVA 1287

Qy 374 SVPLSPATNMSAHRGCVGVSVNGKRLDLTYSFLSGQIGQCQYDSSPCERQPCQHGATC 433
Db 1288 SLRQAPQN-GTSPHGICIRNLYINSELQDFQKVPMTGILPGC---EPCHKVCAGH-TC 1342

Qy 434 MPAGEYFQCLCRDGFKGDLCEHEEN-PCQLRBPCLHGGTCQ-----GTRCICLPFGSGP 487
Db 1343 QPSSQAGFTCECQEGWGMPLCDORTNDPC-LGNKCVH-GTCLPINAFSYSCKLEHGGV 1400

Qy 488 RCOQ 491
Db 1401 LCDE 1404

RESULT 11
US-09-540-153-2
; Sequence 2, Application US/09540153
; Patent No. 6270995
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; CURRENT APPLICATION NUMBER: US/09/540,153
; CURRENT FILING DATE: 2000-03-31
; PRIOR FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: human
US-09-540-153-2

Query Match 10.4%; Score 397.5; DB 3; Length 1525;
Best Local Similarity 29.7%; Pred. No. 1.3e-24;
Matches 108; Conservative 58; Mismatches 141; Indels 57; Gaps 18;

Qy 158 HC-----PTCRDRPCQNGQCHDSESSSYVCVPAGFTGSRCEHSQAL-----HCHPEAC 207
Db 1068 HCDIDFDDCQDNKCKNGAHCCTDA-VNGYTCICPEGYSGLCFCEFPWVLPRTSPCDNFD 1126

Qy 208 GPDATC---VNRPDGRGYTCRHLGRSGLRCEGVTTPSLSGAGSYLALPAL-----T 259
Db 1127 QNGAQCIVRINEP-----ICQCLPGYQGEKELVSV---NFINKESYLQIPSAKVRPQT 1178

Qy 260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEFVSLAMVGGHLEFRYELGSLA-VLRS 318
Db 1179 NITLIQIATDED-----SGILLYKGDK-----DHIAVELYRGRVRSYDTGSHPASAIYS 1227

Qy 319 AEPLALGRWHRVSAERLNKQGLSVNGRVPVLRSSPKSQGLNHLTLILYLGVE-----P 373
Db 1228 VETINDGNFHIIVELLALDQSLSVSDGNGPKIITNLSKQSTLNFDSPLYYGGMPGKSNVA 1287

Qy 374 SVPLSPATNMSAHRGCVGVSVNGKRLDLTYSFLSGQIGQCQYDSSPCERQPCQHGATC 433
Db 1288 SLRQAPQN-GTSPHGICIRNLYINSELQDFQKVPMTGILPGC---EPCHKVCAGH-TC 1342

Qy 434 MPAGEYFQCLCRDGFKGDLCEHEEN-PCQLRBPCLHGGTCQ-----GTRCICLPFGSGP 487
Db 1343 QPSSQAGFTCECQEGWGMPLCDORTNDPC-LGNKCVH-GTCLPINAFSYSCKLEHGGV 1400

Qy 488 RCOQ 491
Db 1401 LCDE 1404

```
Db 1401 LCDE 1404

RESULT 12
US-09-312-283C-396
; Sequence 396, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated from Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c2
; CURRENT APPLICATION NUMBER: US/09/312,283C
; NUMBER OF SEQ ID NOS: 425
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 396
; LENGTH: 1529
; TYPE: PRT
; ORGANISM: Mouse
US-09-312-283C-396

Query Match 10.3%; Score 394.5; DB 4; Length 1529;
Best Local Similarity 30.2%; Pred. No. 2.4e-24;
Matches 110; Conservative 54; Mismatches 143; Indels 57; Gaps 18;

QY 158 HC-----PTCRDRPCQNGGCHDSSESSYVCVPAGTGRCHSQAAL-----HCHPEAC 207
Db 1072 HCDIDPDCQDNKCKNGAHTDA-VNGYTCVCEGSGLFCEFSPPMVLRTSPCDNFD 1130
QY 208 GPDATC---VNRDGRGYTCRCHLGRSLRCEGVTVTTPSLGAGSYLALPAL-----T 259
Db 1131 QNGAQCIIRVNEP-----ICQCLPGYLGEKCEKLVSVI---LVNKESYLQIPSAKVRPQT 1182
QY 260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSLA-VLRS 318
Db 1183 NITLQIATDED-----SGILLYKGDK-----DHIAVELYRGVRASVDTGSHPASAIYS 1231
QY 319 AEPALGRWHRVSAERLNDKGSRLVNGRPRVLRSSPGKSQLNLHTLLYLGGVE-----P 373
Db 1232 VETINDGNFHVILLTDLSSLSVSDGSPKIIITNLKOSTLNFDSPLYVGMPPGKNVA 1291
QY 374 SVPLSPATNMSAHRFCGVCSEVSVNGKRLDITYSFLGSGQIGQCYDSSPCERQPCQHGATC 433
Db 1292 SLRQAPGQN-GTSFHGCIIRNLYINSELQDFRKVPMTGILPGC---EPCHKVCAHG-TC 1346
QY 434 MPAGEYEFQCLCRDGFKGLCEHEEN-PCOLREPCPLHGTCQ-----GTRCLCLPGFSGP 487
Db 1347 QPSSQSGFTCECEBGMWGLCDQRTNDPC-LGNKCVH-GTCLPINAFAFSYCKCLEHGGV 1404
QY 488 RCOQ 491
Db 1405 LCDE 1408

RESULT 13
US-09-312-283C-183
; Sequence 183, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated from Skin Cells
; TITLE OF INVENTION: and Methods for Their Use

Db 1401 LCDE 1404

FILE REFERENCE: 11000.1011c2
; CURRENT APPLICATION NUMBER: US/09/312,283C
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 425
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 183
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Mouse
US-09-312-283C-183

Query Match 10.3%; Score 394; DB 4; Length 716;
Best Local Similarity 28.3%; Pred. No. 8.8e-25;
Matches 126; Conservative 59; Mismatches 150; Indels 100; Gaps 23;

QY 158 HC-----PTCRDRPCQNGGCHDSSESSYVCVPAGTGRCHSQAAL-----HCHPEAC 207
Db 315 HCDIDPDCQDNKCKNGAHTDA-VNGYTCVCEGSGLFCEFSPPMVLRTSPCDNFD 373
QY 208 GPDATC---VNRDGRGYTCRCHLGRSLRCEGVTVTTPSLGAGSYLALPAL-----T 259
Db 374 QNGAQCIIRVNEP-----ICQCLPGYLGEKCEKLVSVI---LVNKESYLQIPSAKVRPQT 426
QY 260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSLA-VLRS 318
Db 427 NITLQIATDED-----SGILLYKGDK-----DHIAVESIEG-IRASYDTGSHPASAIYS 474
QY 319 AEPALGRWHRVSAERLNDKGSRLVNGRPRVLRSSPGKSQLNLHTLLYLGGVE-----P 373
Db 475 VETINDGNFHVILLTDLSSLSVSDGSPKIIITNLKOSTLNFDSPLYVGMPPGKNVA 534
QY 374 SVPLSPATNMSAHRFCGVCSEVSVNGKRLDITYSFLGSGQIGQCYDSSPCERQPCQHGATC 433
Db 535 SLRQAPGQN-GTSFHGCIIRNLYINSELQDFRKVPMTGILPGC---EPCHKVCAHG-TC 589
QY 434 MPAGEYEFQCLCRDGFKGLCEHEEN-PCOLREPCPLHGTCQ-----GTRCLCLPGFSGP 487
Db 590 QPSSQSGFTCECEBGMWGLCDQRTNDPC-LGNKCVH-GTCLPINAFAFSYCKCLEHGGV 647
QY 488 RCOQSGHGAIESDWHLEGGSGNDAPQYGAFFDDGFLAPPG-HVFSRSLPEVPETIEL 546
Db 648 LCDE-----EEDLNFPLPGDQVQAR-----667
QY 547 EVRTSTASGILLW--QGEVEGEAGQ 569
Db 668 EVQALNARAALLMQQWIHRGQLTQ 692

RESULT 14
US-09-182-024A-2
; Sequence 2, Application US/09182024A
; Patent No. 6342370
; GENERAL INFORMATION:
; APPLICANT: Connolly, Timothy
; APPLICANT: Rajput, Bhuan
; TITLE OF INVENTION: Human Slit Polypeptide and Polynucleotides Encoding
; TITLE OF INVENTION: Same
; FILE REFERENCE: 640100-271
; CURRENT APPLICATION NUMBER: US/09/182,024A
; CURRENT FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 60/063,946
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/096,420
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1523
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-182-024A-2

Query Match 10.2%; Score 391.5; DB 4; Length 1523;
```


Search completed: March 9, 2004, 17:23:32
Job time : 29.6355 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:11:22 ; Search time 85.8465 Seconds
(without alignments)
2320.373 Million cell updates/sec

Title: US-10-006-011A-3

Perfect score: 3825

Sequence: 1 EIKITRPDSADGMLLYNQ.....QPLDLQHRAGANTRPCFS 705

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB ID	Description
1	3825	100.0	4391	6	AAE34390	Human per
2	3815	99.7	4393	4	AAE31889	Amino aci
3	3795	99.2	4436	4	ABG23365	Novel hum
4	1030	26.9	195	4	AAE31890	Amino aci
5	901.5	23.6	1931	6	ABU52400	Human GPC
6	891.5	23.3	2053	7	ADC39154	Novel hum
7	849.5	22.2	2143	7	ADC39164	Novel hum
8	840	22.0	671	6	ABP58231	Human cel
9	840	22.0	1009	7	ADP28105	Human NTR
10	832	21.8	819	5	ABP72291	Rat prote
11	812.5	21.2	4072	4	ABP63614	Drosophil
12	664.5	17.4	1298	4	ABP61948	Drosophil
13	640.5	16.7	492	2	AAW26609	Human agr
14	635.5	16.6	1566	7	ADC39166	Novel hum
15	628	16.4	463	7	ADP64904	Human pro
16	625.5	16.4	1565	7	ADC39156	Novel hum
17	605	15.8	432	4	AAU18102	Novel hum
18	605	15.8	432	4	AAU16938	Human nov
19	605	15.8	432	4	ABE10233	Human cdn
20	605	15.8	432	4	AAU19905	Novel hum
21	605	15.8	432	5	ABJ05729	Novel hum
22	605	15.8	432	5	ABP66820	Human pol
23	587.5	15.4	416	2	AAV73993	Human pro
24	576.5	15.1	1039	4	ABP63124	Drosophil
25	556	14.5	1741	5	ABP43859	Human mRN

26	543.5	14.2	375	4	AAU07421	Novel hum
27	441	11.5	3313	7	ADE54405	Rat Prote
28	441	11.5	3313	7	ADE54409	Rat Prote
29	420.5	11.0	1712	4	AAU28190	Novel hum
30	412.5	10.8	1327	4	AAW78871	Human pro
31	412	10.8	1522	4	AAW78855	Human pro
32	407	10.6	1508	2	AAW96706	Protein s
33	407	10.6	1508	2	AAW96706	Protein s
34	407	10.6	1508	2	AAW96706	Protein s
35	407	10.6	1534	2	AAW46966	Amino aci
36	407	10.6	1534	2	AAW27144	Human sli
37	407	10.6	1534	2	AAW96707	Protein s
38	407	10.6	1534	2	AAW96707	Protein s
39	407	10.6	1534	2	AAW96707	Protein s
40	401.5	10.5	1514	7	ADD49063	Human NOV
41	398	10.4	1438	5	AAE17600	Human ext
42	397.5	10.4	1503	2	AAW27142	Human sli
43	397.5	10.4	1503	2	AAW96701	Slit-like
44	397.5	10.4	1525	2	AAW17499	Human sli
45	397.5	10.4	1525	7	ADB85322	Human ner

ALIGNMENTS

RESULT 1

AAE34390
ID AAE34390 standard; protein; 4391 AA.

XX AC AAE34390;

XX DT 14-MAY-2003 (first entry)

XX DE Human perlecan protein.

XX DX Human; diagnosis; osteoarthritis; rheumatoid arthritis; perlecan.

XX OS Homo sapiens.

XX PN WO200295415-A2.

XX PD 28-NOV-2002.

XX PF 22-MAY-2002; 2002WO-EP005612.

XX PR 23-MAY-2001; 2001GB-00013626.

XX PA (OSTE-) OSTEOMETER BIO TECH AS.

XX PI Christgau S, Henriksen DB, Cloos PAC;

XX DR WPI; 2003-140389/13.

XX PT An assay for the diagnosis or assessment of the severity of osteoarthritis or rheumatoid arthritis comprising detecting an isomerized or optically inverted protein in a sample.

XX PS Disclosure; Page 46-67; 106pp; English.

XX CC The invention relates to an assay for the diagnosis or assessment of the severity of osteoarthritis or rheumatoid arthritis. The assay involves measuring (in a biological sample) the amount or presence of an isomerized or optically inverted protein or one or more isomerized or optically inverted fragments from proteins such as perlecan, biglycan, decorin, fibrillin-1 or procadherin. The assay is useful for the diagnosis or assessment of the severity of osteoarthritis or rheumatoid arthritis. The present sequence is human perlecan protein

XX SQ Sequence 4391 AA;

Query Match 100.0%; Score 3825; DB 6; Length 4391;
Best Local Similarity 100.0%; Pred. No. 2.6e-233;
Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	EIKITFRPDSADGMLLYNGQKRVFGSPNTLNAROPDFTSFGLVGRBEPFRDAGSGWATI	60
Db	3687	EIKITFRPDSADGMLLYNGQKRVFGSPNTLNAROPDFTSFGLVGRBEPFRDAGSGWATI	3746
Qy	61	RHPTPLALGHFTHTVTLRSLTQGSLLVGDLPVNGTSGKFGQGLDNLNEELYLGYPDYGA	120
Db	3747	RHPTPLALGHFTHTVTLRSLTQGSLLVGDLPVNGTSGKFGQGLDNLNEELYLGYPDYGA	3806
Qy	121	IPKAGLSGFGICVRELRIQGEETVFHFDLNLNTANGISHCPTCRDRPQNGQCHDSSSS	180
Db	3807	IPKAGLSGFGICVRELRIQGEETVFHFDLNLNTANGISHCPTCRDRPQNGQCHDSSSS	3866
Qy	181	YVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCESGVT	240
Db	3867	YVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCESGVT	3926
Qy	241	VTPPSLSGAGSYLALPALTNTNHEHLRDVDFPKPLAPDGVLLFSGGKSGPVEDFVSLAMVG	300
Db	3927	VTPPSLSGAGSYLALPALTNTNHEHLRDVDFPKPLAPDGVLLFSGGKSGPVEDFVSLAMVG	3986
Qy	301	GHLEFRYELSGGLAVLRSAEPLALGRHWRVSAERLNKDGSLRVNGGRPVTLRSSPGKSGQL	360
Db	3987	GHLEFRYELSGGLAVLRSAEPLALGRHWRVSAERLNKDGSLRVNGGRPVTLRSSPGKSGQL	4046
Qy	361	NLHTLLYLGVEPVSPLSPATNMSAHFPGCVGEVSVNGKRLDLITYSLFGSGQIGQCVDSS	420
Db	4047	NLHTLLYLGVEPVSPLSPATNMSAHFPGCVGEVSVNGKRLDLITYSLFGSGQIGQCVDSS	4106
Qy	421	PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQAREPCLIHGTCQGTCLC	480
Db	4107	PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQAREPCLIHGTCQGTCLC	4166
Qy	481	LPFGSPRCQCGSHGIAESDWHLEGSGGNDAPQYGAIFHDDGFLAPPGHVFPSRSUPEV	540
Db	4167	LPFGSPRCQCGSHGIAESDWHLEGSGGNDAPQYGAIFHDDGFLAPPGHVFPSRSUPEV	4226
Qy	541	PETIELEVRTSTAGLLLWQVEVGEAGQKDFISLGLQDHLVFRYQLGSGEARLVSED	600
Db	4227	PETIELEVRTSTAGLLLWQVEVGEAGQKDFISLGLQDHLVFRYQLGSGEARLVSED	4286
Qy	601	PINDGEHWRVTALREGRGSIQVDGBELVSGRSPGPNNAVNAKGSVYIGAPDVATLTGG	660
Db	4287	PINDGEHWRVTALREGRGSIQVDGBELVSGRSPGPNNAVNAKGSVYIGAPDVATLTGG	4346
Qy	661	RFSSGTCGVNVLVHSARGAPPPQPDLDLQHAQAGANTRPCS	705
Db	4347	RFSSGTCGVNVLVHSARGAPPPQPDLDLQHAQAGANTRPCS	4391

RESIT.T 2

RESULT 2
AAB31889

AA331889
ID AAB31889 standard; protein: 4393 AA.

XX
XX
COOTCIVV DT

AC AAB31889;

XXXXXX

DT 15-MAY-2001 (first entry)

[illegible]

DE Amino acid sequence of a human protein.

1. **Introduction**

2. **Background**

3. **Methods**

4. **Results**

5. **Conclusion**

6. **References**

7. **Appendix**

8. **Supplementary Materials**

9. **Tables**

10. **Figures**

11. **Tables**

12. **Figures**

13. **Tables**

14. **Figures**

15. **Tables**

16. **Figures**

17. **Tables**

18. **Figures**

19. **Tables**

20. **Figures**

21. **Tables**

22. **Figures**

23. **Tables**

24. **Figures**

25. **Tables**

26. **Figures**

27. **Tables**

28. **Figures**

29. **Tables**

30. **Figures**

31. **Tables**

32. **Figures**

33. **Tables**

34. **Figures**

35. **Tables**

36. **Figures**

37. **Tables**

38. **Figures**

39. **Tables**

40. **Figures**

41. **Tables**

42. **Figures**

43. **Tables**

44. **Figures**

45. **Tables**

46. **Figures**

47. **Tables**

48. **Figures**

49. **Tables**

50. **Figures**

51. **Tables**

52. **Figures**

53. **Tables**

54. **Figures**

55. **Tables**

56. **Figures**

57. **Tables**

58. **Figures**

59. **Tables**

60. **Figures**

61. **Tables**

62. **Figures**

63. **Tables**

64. **Figures**

65. **Tables**

66. **Figures**

67. **Tables**

68. **Figures**

69. **Tables**

70. **Figures**

71. **Tables**

72. **Figures**

73. **Tables**

74. **Figures**

75. **Tables**

76. **Figures**

77. **Tables**

78. **Figures**

79. **Tables**

80. **Figures**

81. **Tables**

82. **Figures**

83. **Tables**

84. **Figures**

85. **Tables**

86. **Figures**

87. **Tables**

88. **Figures**

89. **Tables**

90. **Figures**

91. **Tables**

92. **Figures**

93. **Tables**

94. **Figures**

95. **Tables**

96. **Figures**

97. **Tables**

98. **Figures**

99. **Tables**

100. **Figures**

101. **Tables**

102. **Figures**

103. **Tables**

104. **Figures**

105. **Tables**

106. **Figures**

107. **Tables**

108. **Figures**

109. **Tables**

110. **Figures**

111. **Tables**

112. **Figures**

113. **Tables**

114. **Figures**

115. **Tables**

116. **Figures**

117. **Tables**

118. **Figures**

119. **Tables**

120. **Figures**

121. **Tables**

122. **Figures**

123. **Tables**

124. **Figures**

125. **Tables**

126. **Figures**

127. **Tables**

128. **Figures**

129. **Tables**

130. **Figures**

131. **Tables**

132. **Figures**

133. **Tables**

134. **Figures**

135. **Tables**

136. **Figures**

137. **Tables**

138. **Figures**

139. **Tables**

140. **Figures**

141. **Tables**

142. **Figures**

143. **Tables**

144. **Figures**

145. **Tables**

146. **Figures**

147. **Tables**

148. **Figures**

149. **Tables**

150. **Figures**

151. **Tables**

152. **Figures**

153. **Tables**

154. **Figures**

155. **Tables**

156. **Figures**

157. **Tables**

158. **Figures**

159. **Tables**

160. **Figures**

161. **Tables**

162. **Figures**

163. **Tables**

164. **Figures**

165. **Tables**

166. **Figures**

167. **Tables**

168. **Figures**

169. **Tables**

170. **Figures**

171. **Tables**

172. **Figures**

173. **Tables**

174. **Figures**

175. **Tables**

176. **Figures**

177. **Tables**

178. **Figures**

179. **Tables**

180. **Figures**

181. **Tables**

182. **Figures**

183. **Tables**

184. **Figures**

185. **Tables**

186. **Figures**

187. **Tables**

188. **Figures**

189. **Tables**

190. **Figures**

191. **Tables**

192. **Figures**

193. **Tables**

194. **Figures**

195. **Tables**

196. **Figures**

197. **Tables**

198. **Figures**

199. **Tables**

200. **Figures**

201. **Tables**

202. **Figures**

203. **Tables**

204. **Figures**

205. **Tables**

206. **Figures**

207. **Tables**

208. **Figures**

209. **Tables**

210. **Figures**

211. **Tables**

212. **Figures**

213. **Tables**

214. **Figures**

215. **Tables**

216. **Figures**

217. **Tables**

218. **Figures**

219. **Tables**

220. **Figures**

221. **Tables**

222. **Figures**

223. **Tables**

224. **Figures**

225. **Tables**

226. **Figures**

227. **Tables**

228. **Figures**

229. **Tables**

230. **Figures**

231. **Tables**

232. **Figures**

233. **Tables**

234. **Figures**

235. **Tables**

236. **Figures**

237. **Tables**

238. **Figures**

239. **Tables**

240. **Figures**

241. **Tables**

242. **Figures**

243. **Tables**

244. **Figures**

245. **Tables**

246. **Figures**

247. **Tables**

248. **Figures**

249. **Tables**

250. **Figures**

251. **Tables**

252. **Figures**

253. **Tables**

254. **Figures**

255. **Tables**

256. **Figures**

257. **Tables**

258. **Figures**

259. **Tables**

260. **Figures**

261. **Tables**

262. **Figures**

263. **Tables**

264. **Figures**

265. **Tables**

266. **Figures**

267. **Tables**

268. **Figures**

269. **Tables**

270. **Figures**

271. **Tables**

272

KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;

KW ganglioside GM2 activator; saposin B; degenerative disease;

KW neurological
KW Alzheimer's dKW Alzheimer's dis
rheumatoid polyvKW
XX
rheumatoid po

XX
OS
Homo sapiens.

XX
Homo sapiens.

15-JUL-1999; 99PR-00009372.
(INMR) BIOMERIEUX STELHYS.
Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
WPI; 2001-159475/16.
Detecting, preventing and treating degenerative, neurological and autoimmune diseases, particularly multiple sclerosis, using specified polypeptides or related nucleic acid or ligand.
Claim 1; Page 138-152; 203pp; French.
The present sequence represents a human protein, which is used in the method of the invention. The specification describes a method which uses at least one polypeptide or polynucleotide sequence belonging to the p15can, precursor of the retinol-binding plasma protein, precursor of the ganglioside GM2 activator, calgranulin B or saposin B protein families. The method is used for detecting, preventing or treating a degenerative, neurological and/or auto-immune disease. The polynucleotides and polypeptides are used for diagnosis, prognosis, prevention and treatment of multiple sclerosis (in its various forms and phases). They may also be useful in cases of e.g. Alzheimer's and Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid polyarthritis and lupus erythematosus, including use as vaccines and in gene therapy (expression of sense or antisense sequences). They can also be used to assess efficacy of potential therapeutic agents, particularly compounds that reduce or inhibit toxicity towards glial cells
Sequence 4393 AA;

QY 541 PETIELEVRTASGLLLWQGVGEAGQKDFISLGLQDHLVFRYQLGSGEARLVSED 600
 DB 4229 PETIELEVRTASGLLLWQGVGEAGQKDFISLGLQDHLVFRYQLGSGEARLVSED 4288
 QY 601 PINDGEWHRTALREGRGSIQVDGELVSGRPGPNVAVNAKGSVYIGAPDVATLTGG 660
 DB 4289 PINDGEWHRTALREGRGSIQVDGELVSGRPGPNVAVNAKGSVYIGAPDVATLTGG 4348
 QY 661 RFSSGITGCVKLVLSHARPAGPPQDLOHRAQAGANTRPCPS 705
 DB 4349 RFSSGITGCVKLVLSHARPAGPPQDLOHRAQAGANTRPCPS 4393

RESULT 3

ABG23265
 ID ABG23265 standard; protein; 4436 AA.
 AC ABG23265;
 XX
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #23256.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 FN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSRQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS87452.
 XX

XX
 PS New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 53624; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 4436 AA;

Query Match 99.2%; Score 3795; DB 4; Length 4436;
 Best Local Similarity 98.2%; Pred. No. 2.1e-231;
 Matches 701; Conservative 1; Mismatches 2; Indels 10; Gaps 1;

QY 1 EIKITFRPDSADGMLLVNGOKRVPGSPNTLANRPDFTISGLVGRPFRRDAGSGMATI 60
 DB 3722 EIKITFRPDSADGMLLVNGOKRVPGSPNTLANRPDFTISGLVGRPFRRDAGSGMATI 3781
 QY 61 RHPTPLALGHFHTVTLRLSLTQGSLLIVGDLAPVNGTSQGFQGLDLNEELVLYGYPDYGA 120
 DB 3782 RHPTPLALGHFHTVTLRLSLTQGSLLIVGDLAPVNGTSQGFQGLDLNEELVLYGYPDYGA 3841
 QY 121 IPKAGLSGGFICVREIRIQGEEIVFHDNLNLTAGHISHCPTCRDRPCQNGQCHDSSESS 180
 DB 3842 IPKAGLSGGFICVREIRIQGEEIVFHDNLNLTAGHISHCPTCRDRPCQNGQCHDSSESS 3901
 QY 181 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVRPDGRGYTCRCHLGRSGLRCEEGVT 240
 DB 3902 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVRPDGRGYTCRCHLGRSGLRCEEGVT 3961
 QY 241 VTTPLSGAGSYLALPALTNTHTHELRDLDFEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
 DB 3962 VTTPLSGAGSYLALPALTNTHTHELRDLDFEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 4021
 QY 301 GHLEFRYELAGSLAVLRSAEPLALGRWHRYSAERLNKDGSLRVNGRPPVLRSSPGKSOGL 360
 DB 4022 GHLEFRYELAGSLAVLRSAEPLALGRWHRYSAERLNKDGSLRVNGRPPVLRSSPGKSOGL 4081
 QY 361 NHTLLYLVGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLYTSLGSGIGQCYDSS 420
 DB 4082 NHTLLYLVGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLYTSLGSGIGQCYDSS 4141
 QY 421 PCERQPCOHGATCMPAGEYEFQCLCRDGFKGDICEEENPCOLREPCLHGGTCQGTCLC 480
 DB 4142 PCERQPCOHGATCMPAGEYEFQCLCRDGFKGDICEEENPCOLREPCLHGGTCQGTCLC 4201
 QY 481 LPFGSPRCQGGHGAESDMHLEGGSGNDAPQGYGAYFHDGFLAFPGHVSRSLSPEV 540
 DB 4202 LPFGSPRCQGGHGAESDMHLEGGSGNDAPQGYGAYFHDGFLAFPGHVSRSLSPEV 4261
 QY 541 PETIELEVRTASGLLLWQGVGEAGQKDFISLGLQDHLVFRYQLGSGEARLVSED 590
 DB 4262 PETIELEVRTASGLLLWQGVGEAGQKDFISLGLQDHLVFRYQLGSGEARLVSED 4321
 QY 591 SGEARLVSEDPINDGEWHRTALREGRGSIQVDGELVSGRPGPNVAVNAKGSVYIGG 650
 DB 4322 SGEARLVSEDPINDGEWHRTALREGRGSIQVDGELVSGRPGPNVAVNAKGSVYIGG 4381
 QY 651 APDVATLTGGRFSSGITGCVKLVLSHARPAGPPQDLOHRAQAGANTRPCP 704
 DB 4382 APDVATLTGGRFSSGITGCVKLVLSHARPAGPPQDLOHRAQAGANTRPCP 4435

RESULT 4

AAB31890
 ID AAB31890 standard; protein; 195 AA.
 XX
 AC AAB31890;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of the C-terminal of the human perlecan protein.

XX Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 XX

PN WO200105422-A2.
 XX 25-JAN-2001.
 XX 17-JUL-2000; 2000WO-FR002057.
 PF 15-JUL-1999; 99FR-00009372.
 PR (INNR) BIOMERIEUX STELHYS.
 PA Roeklin D, Kolbe H, Charles M, Marcus C, Santoro L, Perron H;
 PI WPI; 2001-159475/16.
 XX N-PSDB; AAF54728.
 DR Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX Claim 1; Page 152-153; 209pp; French.
 XX The present sequence represents a human polypeptide, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 195 AA;
 Query Match 26.9%; Score 1030; DB 4; Length 195;
 Best Local Similarity 100.0%; Pred. No. 1.3e-57;
 Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 511 DAPQYGVFDDGFLAFPHGVFVSRSLPEVPETIEVLTSTASGLLMQGVGEAGGG 570
 Db 1 DAPQYGVFDDGFLAFPHGVFVSRSLPEVPETIEVLTSTASGLLMQGVGEAGGG 60
 QY 571 KDFISLQDGHVFRYQLGSGEARLVSEDPINDGEWHRVTLRGRGSIQVDGEELVS 630
 Db 61 KDFISLQDGHVFRYQLGSGEARLVSEDPINDGEWHRVTLRGRGSIQVDGEELVS 120
 QY 631 GRSPGFNVAVNAGSVYIGAPDVALTGRTSSGTCVKNLVLSARPPGAPPQPLDL 690
 Db 121 GRSPGFNVAVNAGSVYIGAPDVALTGRTSSGTCVKNLVLSARPPGAPPQPLDL 180
 QY 691 QHRAQAGANTRPCPS 705
 Db 181 QHRAQAGANTRPCPS 195
 RESULT 5
 ABUS2400
 ID ABUS2400 standard; protein; 1931 AA.
 XX AC ABUS2400;
 XX AC ABUS2400;
 XX 03-MAR-2003 (first entry)
 XX Human GPCR related protein NOV40a.
 DE Human; NOVX; G-protein coupled receptor; GPCR; cancer; cytostatic.
 XX Human; NOVX; G-protein coupled receptor; GPCR; cancer; cytostatic.
 XX Homo sapiens.
 OS

XX WO200279398-A2.
 PN 10-OCT-2002.
 XX 08-MAR-2002; 2002WO-US007355.
 XX 08-MAR-2001; 2001US-0274194P.
 XX 08-MAR-2001; 2001US-0274281P.
 XX 08-MAR-2001; 2001US-0274322P.
 XX 09-MAR-2001; 2001US-0274849P.
 XX 13-MAR-2001; 2001US-0275578P.
 XX 13-MAR-2001; 2001US-0275579P.
 XX 13-MAR-2001; 2001US-0275601P.
 XX 14-MAR-2001; 2001US-0276000P.
 XX 16-MAR-2001; 2001US-0276776P.
 XX 19-MAR-2001; 2001US-0276994P.
 XX 20-MAR-2001; 2001US-0277239P.
 XX 20-MAR-2001; 2001US-0277327P.
 XX 20-MAR-2001; 2001US-0277338P.
 XX 21-MAR-2001; 2001US-0277791P.
 XX 22-MAR-2001; 2001US-0277833P.
 XX 23-MAR-2001; 2001US-0278152P.
 XX 26-MAR-2001; 2001US-0278894P.
 XX 27-MAR-2001; 2001US-0278999P.
 XX 27-MAR-2001; 2001US-0279036P.
 XX 30-MAR-2001; 2001US-0280233P.
 XX 02-APR-2001; 2001US-0280802P.
 XX 02-MAY-2001; 2001US-0288052P.
 XX 02-MAY-2001; 2001US-0288086P.
 XX 07-MAY-2001; 2001US-0291766P.
 XX 17-MAY-2001; 2001US-0291766P.
 XX 08-JUN-2001; 2001US-0296693P.
 XX 05-JUL-2001; 2001US-0296856P.
 XX 05-JUL-2001; 2001US-0303230P.
 XX 08-AUG-2001; 2001US-0303237P.
 XX 13-AUG-2001; 2001US-0310913P.
 XX 14-AUG-2001; 2001US-0311978P.
 XX 16-AUG-2001; 2001US-0312919P.
 XX 17-AUG-2001; 2001US-0312916P.
 XX 20-AUG-2001; 2001US-0313182P.
 XX 21-AUG-2001; 2001US-0313626P.
 XX 27-AUG-2001; 2001US-0314018P.
 XX 27-AUG-2001; 2001US-0315227P.
 XX 10-SEP-2001; 2001US-0318403P.
 XX 14-SEP-2001; 2001US-0318510P.
 XX 14-SEP-2001; 2001US-0322296P.
 XX 27-SEP-2001; 2001US-0322360P.
 XX 09-NOV-2001; 2001US-0325378P.
 XX 09-NOV-2001; 2001US-032486P.
 XX 07-MAR-2002; 2002US-00094886.
 XX (CURA-) CURAGEN CORP.
 PA Kekuda R, Tchernev VT, Liu X, Spytek KA, Patturajan M;
 PI Burtess CB, Vernet CM, Li L, Gorman L, Malyankar UM, Boldog FL;
 PI Guo X, Shenoy SG, Radigaru M, Taupier RJ, Miller CE, Casman SJ;
 PI Pena CE, Gangoli EA, Gusev V, Smithson G, Zerhusen BD, Gerlach V;
 PI Pochart Fr, Fernandes ER, Shinkets RA, Rastelli L, Spaderna SK;
 PI Larochele WJ, Zhong M, Khrantsov NV, Voss EZ, Herrmann JL;
 XX WPI; 2003-058423/05.
 DR N-PSDB; ABX70491.
 XX NOVX polypeptides and polynucleotides, useful for treating a syndrome
 XX related to a human disease associated with the NOVX polypeptide e.g.,
 XX cancer.
 XX Claim 1; Page 223; 413pp; English.
 XX The present invention relates to the isolation of novel human
 XX polypeptides referred to as NOVX (NOV1-NOV44), variants of these

CC proteins, and the polynucleotide sequences encoding them. The NOVX
 CC proteins of the invention are G-protein coupled receptor (GPCR) related
 CC proteins. The sequences of the invention are useful in the manufacture of
 CC a medicament for treating a syndrome related to a human disease
 CC associated with the polypeptides e.g. cancer. ABU52311-ABU52408 represent
 CC the human NOVX proteins of the invention
 XX
 SQ Sequence 1931 AA;

Query Match 23.68; Score 901.5; DB 6; Length 1931;
 Best Local Similarity 32.28; Pred. No. 2.5e-48;
 Matches 242; Conservative 93; Mismatches 266; Indels 151; Gaps 22;

QY 2 IKTFPDSADGMLLYNQKRVFSPNTLANRPDIFSLVGRPFRRFDAGSGMATIR 61
 Db 1282 LALEFRALFPOGLLYNGNA-----RKGDFLALALLDGRVQLAFDVGSPAVLT 1330
 QY 62 HPTPLALGHPTVTLARSULTGSLIYVDLAPYNGTSGKFGQLDINEELYLGYPD---Y 118
 Db 1331 SAVFVFGQHRLELSRHRWRRGLTSDGETPVLGESPSGTDGLNLDLTFVGVPEDQAA 1390
 QY 119 GAIPKAGLSGFGTCVRELRIOGEEIVFHDNL-----TAHGISHC--PTCRDRPCQNG 170
 Db 1391 VALERTFVAGLRCICLLDNNQRL---ELGIGPGAATRGSGVGECHGHPCLPNCGG 1447
 QY 171 GQCHDSSESYVCVPAGFTGSECHSQAHLCHPEACGPDATCVARPDGRGTYTCRHLGR 230
 Db 1448 APCQNTLEAGRFHQCPFGPRVPTCADEKS-PCQPNPCHGAACRVLPEG-GAQCCEPLGR 1505
 QY 231 SGLRCBEGVTVTTPSLSGAGSYLALPALTNTHELR-----LDVEFKPLAP 276
 Db 1506 EGTFCQ-----TASGQDGGPFLA-DNFGFSLHLEURLHTFARDLGEKVALEVFVLRGP 1559
 QY 277 DGVLLFSGKSGVEDFVSLAMVGHLEFRYELGSLAVLRSAEPLALGRMHRVSAERLN 336
 Db 1560 SGILLYNGQTKDQKDFVSLARDLREFRYDLGGAIVRSRFPVTLGAWTRVSLERNG 1619
 QY 337 KQSLRVNGSRPVLRSPPCKSQGLNLTLLYLGVPSPVPLSPATNMAHFRGCVGVSV 396
 Db 1620 RKGALRVGDPRLVSGSPVPHVTNLLKEPLYVGGAPDFSKLARAASVSGDFGAIQLVSL 1679
 QY 397 NGKRL-----DLTYSFLSGSIGQCYDSSPCR---OPCQHGATCMPAGEYEFQC 443
 Db 1680 GGRQLLTPEHVLROVDVT-SFAG-----HCTRASGHPCLNGASCVP----- 1720
 QY 444 LCEDGFKGLCHEENPCOLRPECPLHGTCQCTRLCLTPGSGPRCQOQSGHGTAESDMW 503
 Db 1721 -----REAY-----VCLCPGSGFSGPCHCKGL----- 1742
 QY 504 LEGSGNDAPQGYGAYFHDGDFLAPPGHVFSRSLPEVPET-----LELEVRTSTASG 555
 Db 1743 VEKSAG-----DVTDLADGRTFVEYLNATVTESEKALQSNHFLSLRTEATQG 1790
 QY 556 LLLWQGEVGEAGQKDFLSLQDGLVFRYQLGSGEARLVSEDPINDGEHVRVTLRE 615
 Db 1791 LVLWS-----GKATERADYVALAIVDGHQLQLSLNGSQPWLRLSTVPVNTNWLAVVAHRE 1846
 QY 616 GREGIQTVDGEELVSGRSPGNVAVNAGSVYIGAPD--VATLTGGRFPSSGIGTCVKUL 673
 Db 1847 QREGSLQVNEAPVTGSSPLGATQDTCALWGLGLPELPGVLPALPKAYGTGFGVGLRDV 1906
 QY 674 VLHSARPAGPPQPLDLOHRAQAGANTRPCPS 705
 Db 1907 VVGR-----HPLHLLEDAVTKPELRPCPT 1930

RESULT 6
 ADC39154
 ID ADC39154 standard; protein; 2053 AA.
 AC ADC39154;
 XX
 XX
 DT 18-DEC-2003 (first entry)

XX DE Novel human NOVX polypeptide SEQ ID NO: 96.
 XX
 KW antidiabetic; cyrostatic; immunomodulator; anorectic; antilipemic;
 KW nootropic; neuroprotective; immunostimulant; antiparkinsonian; anti-HIV;
 KW antiasthmatic; antiinflammatory; hypotensive; antiarteriosclerotic;
 KW hemostatic; osteopathic; gene therapy.; NOVX; diabetes; cancer;
 KW lymphoma; uterus cancer; prostate cancer; dyslipidemia; anorexia;
 KW wasting disorder; Alzheimer's disease; Parkinson's disorder; cachexia;
 KW cardiomyopathy; AIDS; asthma; Crohn's disease; multiple sclerosis;
 KW hypertension; atherosclerosis; hemophilia; graft-versus-host disease;
 KW Albright hereditary osteodystrophy.
 XX OS Homo sapiens.
 XX WO2003010327-A2.
 XX PD 06-FEB-2003.
 XX PF 02-MAY-2002; 2002WO-US014199.
 XX PR 02-MAY-2001; 2001US-0288063P.
 PR 03-MAY-2001; 2001US-0288395P.
 PR 07-MAY-2001; 2001US-0289087P.
 PR 09-MAY-2001; 2001US-0289817P.
 PR 09-MAY-2001; 2001US-0289818P.
 PR 11-MAY-2001; 2001US-0290194P.
 PR 14-MAY-2001; 2001US-0290753P.
 PR 15-MAY-2001; 2001US-0291181P.
 PR 16-MAY-2001; 2001US-0291243P.
 PR 18-MAY-2001; 2001US-0292001P.
 PR 21-MAY-2001; 2001US-0292374P.
 PR 22-MAY-2001; 2001US-0292587P.
 PR 23-MAY-2001; 2001US-0293107P.
 PR 25-MAY-2001; 2001US-0293747P.
 PR 29-MAY-2001; 2001US-0294109P.
 PR 29-MAY-2001; 2001US-0294110P.
 PR 30-MAY-2001; 2001US-0294434P.
 PR 31-MAY-2001; 2001US-0294827P.
 PR 12-JUL-2001; 2001US-0304879P.
 PR 31-JUL-2001; 2001US-0308901P.
 PR 14-AUG-2001; 2001US-0312270P.
 PR 17-AUG-2001; 2001US-0313416P.
 PR 10-SEP-2001; 2001US-0318463P.
 PR 27-SEP-2001; 2001US-0325683P.
 PR 18-OCT-2001; 2001US-0330292P.
 PR 28-NOV-2001; 2001US-0333873P.
 PR 03-DEC-2001; 2001US-0336909P.
 PR 03-DEC-2001; 2001US-0337552P.
 PR 21-FEB-2002; 2002US-0359245P.
 PR 01-MAY-2002; 2002US-00136826.
 XX (CURA-) CURAGEN CORP.

Miller CE, Kekuda R, Malyankar UM, Li L, Pena CEA, Spytek KA;
 Gorman L, Guo X, Fernandes ER, Smithson G, Stone DV, Zerkhusen BD;
 Patturajan M, Anderson DW, Mezes PS, Feyman JA, Macdougall JR;
 Padigaru M, Rastelli L, Shenoy SG, Gerlach VL, Shinkets RA, Zhong M;
 Edinger SR, Ellerman K;
 WPI; 2003-239445/23.
 N-PSDB; ADC39153.
 PT New NOVX polypeptides and polynucleotides, useful in gene therapy,
 PT particularly for treating or preventing a syndrome associated with a
 PT human disease e.g. diabetes, obesity, cancer, Alzheimer's disease,
 PT hypertension or hemophilia.
 XX Claim 1; SEQ ID NO 96; 748pp; English.

The invention relates to new isolated NOVX polypeptides, the genes
 encoding them or sequences having at least 95% identity to the amino acid
 or nucleotide sequences. The NOVX polypeptide is useful as a therapeutic,

New NOVX polypeptides and polynucleotides, useful in gene therapy, particularly for treating or preventing a syndrome associated with a human disease e.g. diabetes, obesity, cancer, Alzheimer's disease, hypertension or hemophilia.

Claim 1; SEQ ID NO 106; 748pp; English.

The invention relates to new isolated NOVX polypeptides, the genes encoding them or sequences having at least 95% identity to the amino acid or nucleotide sequences. The NOVX polypeptide is useful as a therapeutic, particularly in the manufacture of a medicament for treating a syndrome associated with a human disease, which includes a pathology associated with NOVX polypeptide. The NOVX polypeptide is particularly useful for treating, preventing or alleviating pathology associated with NOVX polypeptide in a mammal, e.g. a human. The NOVX nucleic acid and polypeptide are especially useful for treating or preventing e.g. diabetes, obesity, cancers (e.g. lymphoma, uterus cancer or prostate cancer), dyslipidemias, anorexia, wasting disorders, Alzheimer's disease, Parkinson's disorder, cachexia, cardiomyopathy, AIDS, asthma, Crohn's disease, multiple sclerosis, hypertension, atherosclerosis, hemophilia, graft-versus-host disease or Albritch hereditary osteodystrophy. The DNA encoding the protein is useful in gene therapy for treating the above conditions. These are also useful in developing powerful assay system for functional analysis of various human disorders, as well as in diagnostic applications. This sequence represents one of the NOVX proteins of the invention.

Sequence 2143 AA;

Query Match 22.2%; Score 849.5; DB 7; Length 2143;
Best Local Similarity 29.0%; Pred. No. 5.7e-45;
Matches 246; Conservative 95; Mismatches 259; Indels 247; Gaps 26;
2 IKITRPSADGMLYNGKRVGSPNTIANRQDFISGLVGGPRFDFAGSGMAR 61
1400 LALEFRALFPOGLLYNGA-----RGKDFLALLDGRVQLRFDGSGPAVLT 1448
62 HPTPLAIGHFTVTLRLSTLQSLVGLDAPYNTSQKFQGLDNEELYGGYD---Y 118
1449 SAVFVPGQWHRLSRHWRRLTSLVDGETFVLGSPSGTDLNLDTLFVGVPDQAA 1508
119 GAIPKAGLSGFGICVRELRIQGEIVFHDLM-----TAGHSIC--PTCRDRPCQNG 170
1509 VALERTFVGAGLRGICRLDNNQRL---ELGIGPGAATRGSGVCEGDHFLNPNCHGG 1565
171 GOCHDSSESSVYVCVCPAGTSGRCHSQALHCHPCACPDATCNRPDGRGYTCRHLGR 230
1566 AFQCNLEAGRHCQCPGRGVGPTCADEKS--PCQPNCHGAAPCRVLPREG-GAQCECPILGR 1623
231 SGLRCBEGVTVTTPSLGAGSYALPALNTNTHLELR-----LDVDFKPLAP 276
1624 EGTFCQ-----TASGDGSGPPLA-DFNGFSLHRLGLHTIARDLGEKMALEAVFLARGP 1677
277 DVLLFSGKSGPVDVFLAMVGHLFRVELSGSLAVLSAEPLALGRHVRVSAERLN 336
1678 SGLLYNGKRTDGGDFVSLALDRRLFRYDLGKGAIVRSRPFVTLGAWTRVSLERNG 1737
337 KDGSLRVNGRPVLRSSPKSGQ-----LNHTLLYLGGVPSVPLSPATWMSAHFGCV 391
1738 RKGLRVGSGRVLGESP-KSRKVPHTVNLKEPLVYGGAPDFSKLARAASVSGFPGAI 1796
392 GEVSVNGKRL-----DLTVSFLSGSGIGQCYDSSPCR---OPCQHGATCMPAGE 438
1797 QLVSLGGFQLLTPHVLVLRQVDVT-SFAG-----HPCTRASGHFCLNGASCVP--- 1842
439 YEFQCLCRDGFKGLDCEHEENPCQLREPLCHGGTCQGRCLCLFPGSPRCCQSGHGIA 498
1843 -----REAY-----VCLCPGFGSPHCERGL-----1864
499 ESDMHEGSGNDAPGQYAYFHDGGLAFPGHVRFSRLPEVPT-----ILEVRT 550
1865 -----VEKSAG-----DVDTLAFDGRTFVEVLNAVTESEKALQSNHFLSLRT 1907

QY 551 STASGLLLWQGVVEGEAGQGHDFISLGLQDGHVFLRYQLGSGEARLVSEDPINDEWHRV 610
Db 1908 EATQGLVLS-----GKATERADYVALAIVDGHQLSYNLGSPVVLARSTVFNTRMLRV 1963
QY 611 TALREGREGSIQVDEELVSGSPCPNVAIVAKGSVIG-----GAPVAITLG--- 659
Db 1964 VAHREOREGSLQVNEAPVTGSSPLGATQDLDGALWLGECFGTREGCPRVSDIRGTD 2023
QY 660 -----GR----- 661
Db 2024 TFPQPLSQGACEPAGWALFSCSPSCQCGAGRRVRCPPAGHLLVGVPISTESQP 2083
QY 662 -----FSSGITGVKMLVLSARPGAPPPQDLOHRAQAGA 698
Db 2084 GDSHLCRCRGLFELPVGPAFLPKAYGTGFGVGLRDVVVGR-----HPLHLLLEDAVTKP 2135
QY 699 NTRPCPS 705
Db 2136 ELRCPCT 2142
RESULT 8
ID ABP58231 standard; protein; 671 AA.
XX AC ABP58231;
XX DT 31-MAR-2003 (first entry)
XX DE Human cell adhesion and extracellular matrix protein 8.
XX KW Cell adhesion and extracellular matrix protein 8; CADECM-8; human;
KW anti-HIV; virucide; antiallergic; antiinflammatory; antianaemic;
KW antiParkinsonian; nootropic; anticonvulsant; antiinfertility;
KW antiarteriosclerotic; antiasthmatic; immunosuppressive; antithyroid;
KW cytosatic; hepatotropic; dermatological; osteopathic; nephrotropic;
KW angiot; thymelintic; neuroprotective; osteopathic; antithyroid;
KW antiparasitic; antihelminic; antipruritic; uropathic; ophthalmological;
KW antirheumatic; haemostatic; antibacterial; protozoacide; fungicide;
KW gynaecological; gene therapy.
XX OS Homo sapiens.
XX EN WO200289322-A2.
XX PD 07-NOV-2002.
XX PF 01-MAY-2002; 2002WO-US013874.
XX PR 02-MAY-2001; 2001US-0288290P.
XX PR 21-MAY-2001; 2001US-0292468P.
XX PR 15-JUN-2001; 2001US-0298616P.
XX PR 28-JUN-2001; 2001US-0301672P.
XX PR 04-JAN-2002; 2002US-0345008P.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Yue H, Lee EA, Duggan BM, Thangavelu K, Honchell CD, Ding L,
PI Hillman JL, Baughn MR, Kallick DA, Lee S, Warren BA, Xu Y, Tran UK,
PI Lal PG, Thornton M, Hafalia AJA, Yao MG, Nguyen DB, Gandhi AR,
PI Khan FA, Walhia NK, Griffin JA, Chinn AM, Elliott VS, Ramkumar J;
PI Arvizu CS, Forsythe IJ;
XX WPI; 2003-167112/16.
XX N-PSDB; AB244595.
XX New human cell adhesion and extracellular matrix proteins, useful for
PT diagnosing, treating or preventing autoimmune or inflammatory disorder
PT (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,
PT cancer or hepatitis.
XX Claim 1; Page 157-159; 178pp; English.

QY 647 YTGAPDVATLTGGFSSGTTGCVKLVHSGARPCAPPDPLDQHRQAAGANTRPC 703
 DB 767 YVGMKEIALHNRQYMRGLVGCISHFTLST-----DYHISLVEDAVGKNINTC 816
 RESULT 11
 ABB63614
 ID ABB63614 standard; protein; 4072 AA.
 AC ABB63614;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 DE Drosophila melanogaster polypeptide SEQ ID NO 17634.
 XX
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmacological.
 KW Drosophila melanogaster.
 OS
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 FA (PEKE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL07717.
 DR
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 PS Disclosure; SEQ ID NO 17634; 21pp + Sequence Listing; English.
 CC
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB01840-AB01845), expressed DNA
 CC sequences (AB01840-AB01845) and the encoded proteins (AB01840-AB01845).
 CC (AB01840-AB01845) and the sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 4072 AA;
 Query Match 21.2%; Score 812.5; DB 4; Length 4072;
 Best Local Similarity 30.2%; Pred. No. 2.6e-42;
 Matches 219; Conservative 121; Mismatches 304; Indels 81; Gaps 22;
 QY 3 KITFRPDSADGMLYNGQKRVPSPTNLNRQPDRTSFGLVGRPEFRFDAGSGNATIRH 62
 DB 3394 ELTFRPENGDLFLFNGQTRSG-----DYALSLEKRYAEFRFDGKPMVLRA 3443
 QY 63 PTPALGHFTVTLRLSLTQGSLLVGLDLPVNGTSGQKFGQGLDLNBEYLGGYPDYGAIP 122
 DB 3444 EEPALNENHVTVRVSRFRKRGVYQVDEQHPVAFPTLQQLDIEDLYIGGVNELLP 3503
 QY 123 KAGLSS--GFIGVRELRIGEEIVFDNLNLTATGISHCPTCDRDPCCNGGCHDSSES-S 179
 DB 3504 ADAVSQOVGFVGCISRLTQGRVTLIRKATYKEGITDCRCAQGCQNKGVLESQTEQ 3563
 QY 180 SYVVCVPAGFTGRCBHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSLRCEGV 239

DB 3564 ATTCIQPQWTRDCA-IEGTCTCTPGVCGA-GRCENTEN--DMECLCLNRSRGCRCYNE 3619
 QY 240 TVTTPSLSGAGSYLA---LPALTNTTHHELRLDVFEKPLA-PDGVLLFSGGKGPVEDFVS 295
 DB 3620 ILNEHSLNFKGNSFAAYGTGPKVT---KVNITLSVRPASLEDSVILYTAESLTPSGDILA 3675
 QY 296 LAMVGHLEFRYELGSL--AVLRSAPLALGRWHRVSAERLNKDGSLRVNGRPRVLRSS 353
 DB 3676 LVIRGGHAELLINTAARLDFFVVRSAEPLPLNRWTRIEIRRLGEGILRVGDGPRKAKA 3735
 QY 354 PGKSQLNLHLLTYLGGVEPS-VPLSPATNMSAHPGCVGVSVNGKRLDLTYSLGSGQ 412
 DB 3736 PGSDRILSLKTHLYVGGYDRSTVKVNRDYNITKFGDGCISRLYNFKQKPVNLLADIKDAAN 3795
 QY 413 IQCYDSSPCERQPCOHGATCMPAGEYE-----FQCLCRDGFKGLCEHEENPCQLREP 466
 DB 3796 IQSC-----GETNNIGDESDNEPPVPPTPDVHENELOFYAMAPC-ASDP 3841
 QY 467 CLFGGTCQGTTR----CLCLPGFSGPRCQGGHGAESDWHLEGGSGNDAPQGVAYFHD 522
 DB 3842 CENGSCSEQEDVAVCSGPFSGKHCB-----HLQLG-----FNASFRG 3882
 QY 523 DGLFAPPGHVRSLPEVPTIELEVTSTASGLLMOGVVGRAGQKQFISLGLQGH 582
 DB 3883 DGYVELNRSHFQPALEQSYTSMGI VFTTNKPNGLLFWMGQAGBEYTCQDFIAA VVDGY 3942
 QY 583 LVFRYQLGSGEARLVSED-PINDGEHRTALRSGRSGSIOVDGEELVSGRS-PGPNVAV 640
 DB 3943 VEYSNRLDGEAEVIRNSDIRVDNGERHIVIAKRNTAILLEVD-RMLHSGRTRTSKSM 4001
 QY 641 NAKGSVIYGGAPDVATLTGGRFSSGTTGCVKLVHSGARPCAPPDPLDQHRQAAGANT 700
 DB 4002 KLPGNVYVGGAPDLEVTGFTGFKYKINLNCI--VVVEGTG-----QINLSSAAVNGVNA 4054
 QY 701 RPPCS 705
 DB 4055 NVCPA 4059
 RESULT 12
 ABB61948
 ID ABB61948 standard; protein; 1298 AA.
 AC ABB61948;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 12636.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL06051.
 DR
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell

interactions.

PT XX Disclosure; SEQ ID NO 12636; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from *Drosophila*. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA

CC sequences (ABU01840-ABU16175) and the encoded proteins (ABE57737-

CC ABE72072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO.int/pub/published_pct_sequences

XX

SQ Sequence 1298 AA;

Query Match 17.4%; Score 664.5; DB 4; Length 1298;

Best Local Similarity 22.0%; Pred. No. 1.7e-33;

Matches 221; Conservative 111; Mismatches 295; Indels 379; Gaps 23;

QY 1 EIKITRPDSADGMLLVNGOKRVPFGSPFTNLNRPQDFISGLVGGPFRFDAGSGMATI 60

DB 361 QLRLEFRPSFDGILLISGER-----DDITGDFMALLLNKGFVFWFCGSGVGV 411

QY 61 RHPTPLALGHFHTVTLRLSLTQSLIVGDIAPVNGTSQKFGQLDLNEELVGLGYPDY-G 119

DB 412 RSRETIILNENWSVILYRHRDAMLVNLHGCTKVQGRSNGLFSRITFRFPVFLGGIGNITG 471

QY 120 AIPKAGLSSGFIQVRELRLQGEIVFHDNLTAH-----GISHCPT--CRDRPQC 168

DB 472 LAKRLPLAEGFACIR--REVAVE---HDYKFEHPLGDIVNGFDIODCSTDKCVRYPQC 526

QY 169 NGQCQHDSESSVYVCPAGFTGSRCHESQALHCHPEACGPATCVNRPDGRGYTCRCHL 228

DB 527 HGGKCLPSDOGA-ICLCPIGFVGLCE----- 552

QY 229 GRGLRCCEGVVTTTSLSAGSYLALPALTNTTHELRLDVEFKPLADPGVLLFSGGKGG 288

DB 553 -----LRMD-----LQVPAFNGSSFLRYAPLGDLSALIWLEKLVTLKPEQADGLILYSGPEHR 604

QY 289 PVEDFVSLAMVGHLFRYELGSLVLAESAELALGRWHVSAERLNKDGSLRVNGGRP 348

DB 605 --GDFTALYNDGFVEAFPLDGLSGPALVRSEHSLGQWHTIKISTARLAVLKVDKHQE 662

QY 349 VLRSPPKSGQLNLHTLLYLGGVEPVP-SPATNMSAHFRGCVSVNGKRLDLYSEL 408

DB 663 VLATISSNGFWHLSDQNLVFGVGNVHVDRLPLDKYKPFVFCIQRIDNGHSLGIVAEAL 722

QY 409 GSGQIGQCYDSSPCERQPCQHQATCMPAGE-YEFQCLCRDGFKGLDCEH--EENPCOLRE 465

DB 723 GGSNIGNC--PHACVAPCPGLAECVQMESYB-----CRCSIHNERCKAAEVPPEQLPE 776

QY 466 PCLHGGTFCQTR----- 477

DB 777 LALHKSIVLETKONGEAAKVKVSLGAAKHKSHKRNHLKPTPATSTTTTSTTTTITTEAP 836

QY 478 -----CLCL 481

DB 837 SERTEATAGALSNEEIEDIIIFRLVQQQQQKELKKHQQTITPATSTTSSGFSKAK 896

QY 482 GPFSG----- 486

DB 897 PRLSGKHASKHEHLKPNAAFTKLSRULPHTHYESFTNPDSILTFEDNDWVTSLOQ 956

QY 487 -----PRCQQG----- 492

DB 957 EYGDMAASQVPLAFEDASPGTFRSDNNEDDENAFVDFSLDASDGTTEYCRKQLAQD 1016

QY 493 -----SGHGLAESDWHLEGS-----GGNDAPQCYGAVFHDGFL----- 526

DB 1017 MKRIMSNNAHSHKKAELVQFPQSGQEVGTANEDTSQYSDNDNDELTLTPVMOGEEVK 1076

QY 527 -----APFG-----HV 532

DB 1077 LEQHTSTPTQHTDWSLLKKFDLSAEHQSQVQVRKNFGACFAGSDSYFHTNDADTMSQV 1136

QY 533 FSRSLPEVPETIELEVRTSTASGILLMGVVEVGEAGQKDFISGLQDGHVFRYOLGSG 592

DB 1137 IYSISL-----DLNLRKTHSENGVILWTGRQ-GTTEHDDYLSIGISQGYLHFRYDLGSG 1190

QY 593 EARL-VSEDPINDGEHVRVLTALREGRGSIQVDEELVSGRSGPPNVAVNAKSGSVYIGGA 651

DB 1191 EVDIRFNGTKVSDGLWHRVRAIRNSQSGYLEVDGRKVTTLRAPKRLQLNLTDTGLYVCGM 1250

QY 652 PDVATLTGGRSSGIGTCVKNLVLSHARPAPPPQPLDLOHRAQAG 697

DB 1251 PDVGFTHQRFSGIVGCISEIVLAGEMKLNFDNTLTGTEHNVEIG 1296

RESULT 13

AAW26609

ID AAW26609 standard; protein, 492 AA.

XX AAW26609;

XX 27-JAN-1998 (first entry)

XX Human agrin.

XX Agrin; human; receptor tyrosine kinase; muscle specific kinase; Musk;

XX ligand; muscle atrophy; muscular dystrophy; myopathy; diagnosis;

XX gene therapy.

XX Homo sapiens.

XX Key

XX Location/Qualifiers

XX 1..23

XX /label= Sig_peptide

XX 24..492

XX /note= "Claim 2"

XX 60..492

XX /note= "Claim 2"

XX 76..492

XX /note= "Claim 2"

XX 126..492

XX /note= "Claim 2"

XX 178..492

XX /note= "Claim 2"

XX 187..190

XX /label= Y-insert

XX /note= "sequence absent from rat agrin"

XX 222..492

XX /note= "Claim 2"

XX 260..492

XX /note= "Claim 2"

XX 300..492

XX /note= "Claim 2"

XX 324..331

XX /label= Z-insert

XX /note= "sequence absent from rat agrin"

XX

PN WO9721811-A2.

XX

PD 19-JUN-1997.

XX

PF 13-DEC-1996; 96WO-US020696.

XX

PR 15-DEC-1995; 95US-0008657P.

PR 10-MAY-1996; 96US-00644271.

XX

PA (REGE-) REGENERON PHARM INC.

XX

PI Valenzuela DM, Glass DJ, Bowen DC, Yancopoulos GD;

XX WPI, 1997-332783/30.

DR N-PSDB; AAT90471.
XX Nucleotide sequences encoding human agrin and muscle specific kinase and
PT related receptor - used in diagnosis and treatment of disorder with
PT muscle atrophy.
XX
PS Claim 2; Fig 15; 120pp; English.
XX
CC This polypeptide comprises the amino acid sequence of human agrin deduced
CC from an isolated fetal brain cDNA clone (see AAT90471). Full-length
CC agrin, or active portions of it, can be expressed using a claimed host-
CC vector system. The agrin polypeptides can be used in claimed methods of
CC treating diseases or disorders that affect muscle, especially atrophy
CC resulting from denervation due to nerve trauma, degenerative, metabolic
CC or inflammatory neuropathy, peripheral neuropathy or damage to nerves
CC caused by environmental toxins or drugs. In particular the muscle atrophy
CC is due to motor neuropathy, chronic disease, metabolic stress or
CC nutritional insufficiency, muscular dystrophy syndrome, congenital
CC myopathy or acquired (toxic or inflammatory) myopathy. The polypeptides
CC can also be used in a claimed method of promoting the growth,
CC differentiation or survival of cells that express muscle-specific (MuSK)
CC receptors (see AAW26610-11), and to raise antibodies
XX
SQ Sequence 492 AA;
Query Match 16.7%; Score 640.5; DB 2; Length 492;
Best Local Similarity 33.0%; Pred. No. 1.9e-32;
Matches 183; Conservative 59; Mismatches 173; Indels 139; Gaps 18;
QY 202 CHPEACGPDATCVNRPDGRCVTCHLGRSLRCEEGVTWTPSLSGAGSVLALPALNT 261
Db 27 CQNPCHGAACRVLPFG-GAQCCPLQRETFQC-----TASQDSSGFFLA-DENGFS 79
QY 262 FHELR-----LDVEFKPLAPDGVLLFSGKSGFVEDFVSLAMVGHLEFRY 307
Db 80 HLELRGLHTFARDLCEKMALEWFLARGPSGLLYNGOKTDGKDFVSLALDRRLFRY 139
QY 308 ELGSLAVLRGAELPLAGWHRVGAERLNDKGRVNGRRVLRSSPGKSG-----LNL 362
Db 140 DLGGAIVIRREPVTILGAWTRVSLERNRGRKALRVGDGPVLRGESP-KSKKVPHTVNL 198
QY 363 HTLLYLGVEPSVPLSATNNSAHRGCVGEVSVNGKRL-----DLTVSFLGSQG 412
Db 199 KEPLVVGAPDFSKLARAASVSSGPDGAIQLVSLGGRQLLTPEHVLQRQDVT-SFAG--- 254
QY 413 IQCQVDSFPER---OPCHGATCMPAGEYEFQCLCRDGFKGDCEHEENPCQLREPLH 469
Db 255 -----HPCTRASGHPCLNGASCVP-----REAY- 278
QY 470 GGTCCGTCLCLLPFGSGPRCQGGSHGIAESDWHLEGGSGNDAPGQYGAFHDDGFLAPP 529
Db 279 -----VCLCPGSGPHCEKGL-----VEKSAG-----DVTTLAFD 309
QY 530 GHVFSRSLPEVPET-----TELEVRISTASGLLWQGVGEAGQCKDF 573
Db 310 GRTFVEYLNAVTESELANEIPVEKALQSNHFFELSLRTEATQGLVLS-----GKATERADY 365
QY 574 ISLGLQDGHVFRYQLGSGEARLVEEDPINDGEHVRVTLREGRRGSIQVDGELVSGRS 633
Db 366 VALAIVDGHQLQSVNLGSGPVVLRVTVPNTNMLRVVAHREQSGSLQVGNAPVTGSS 425
QY 634 PGNPNVNAKSGVYIGGAPD--VATLTGGRFSSGTGCKVKNLVLSARPGAPPQPDLDLQ 691
Db 426 PLGATQLDGDGALWGLGLPELPGVPALPKAYGTGFGVGLRDVVGR-----HPLHL 477
QY 692 HRAQAGANTRPCPS 705
Db 478 EDVATKPELRPCPT 491
RESULT 14
ADC39166
ID ADC39166 standard; protein; 1566 AA.

XX ADC39166;
XX AC
XX 18-DEC-2003 (first entry)
DT Novel human NOVX polypeptide SEQ ID NO: 108.
XX DE
XX KW antidiabetic; cyrostatic; immunomodulator; anorectic; antilipemic;
KW nootropic; neuroprotective; immunostimulant; antiparkinsonian; anti-HIV;
KW antiasthmatic; antiinflammatory; hypotensive; antiarteriosclerotic;
KW hemostatic; osteopathic; gene therapy; NOVX; diabetes; obesity; cancer;
KW lymphoma; uterus cancer; prostate cancer; dyslipidemia; anorexia;
KW wasting disorder; Alzheimer's disease; Parkinson's disorder; cachexia;
KW cardiomyopathy; AIDS; asthma; Crohn's disease; multiple sclerosis;
KW hypertension; atherosclerosis; hemophilia; graft-versus-host disease;
KW Albright hereditary osteodystrophy.
XX
XX Homo sapiens.
OS
XX WO2003010327-A2.
XX
XX 06-FEB-2003.
XX
XX 02-MAY-2002; 2002WO-US014199.
XX
XX 02-MAY-2001; 2001US-0288063P.
XX 03-MAY-2001; 2001US-0288395P.
XX 07-MAY-2001; 2001US-0289087P.
XX 09-MAY-2001; 2001US-0289817P.
XX 09-MAY-2001; 2001US-0289818P.
XX 11-MAY-2001; 2001US-0290194P.
XX 14-MAY-2001; 2001US-0290753P.
XX 15-MAY-2001; 2001US-0291181P.
XX 16-MAY-2001; 2001US-0291243P.
XX 18-MAY-2001; 2001US-0292001P.
XX 21-MAY-2001; 2001US-0292374P.
XX 22-MAY-2001; 2001US-0292587P.
XX 23-MAY-2001; 2001US-0293107P.
XX 25-MAY-2001; 2001US-0293747P.
XX 29-MAY-2001; 2001US-0294109P.
XX 29-MAY-2001; 2001US-0294110P.
XX 30-MAY-2001; 2001US-0294434P.
XX 31-MAY-2001; 2001US-0294827P.
XX 12-JUL-2001; 2001US-0304879P.
XX 31-JUL-2001; 2001US-0308901P.
XX 14-AUG-2001; 2001US-0312270P.
XX 17-AUG-2001; 2001US-0313416P.
XX 10-SEP-2001; 2001US-0318463P.
XX 27-SEP-2001; 2001US-0325683P.
XX 18-OCT-2001; 2001US-0330292P.
XX 28-NOV-2001; 2001US-0333873P.
XX 03-DEC-2001; 2001US-0336909P.
XX 03-DEC-2001; 2001US-0337552P.
XX 21-FEB-2002; 2002US-0359245P.
XX 01-MAY-2002; 2002US-00136826.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Miller CE, Kekuda R, Malyankar UM, Li L, Pena CEA, Spytek KA;
XX Gorman L, Guo X, Fernandes ER, Smithson G, Stone DJ, Zerhusen BD;
XX Patturajan M, Anderson DW, Mezes PS, Peyman JA, Macdougall JR;
XX Padigaru M, Rastelli L, Shenoy SG, Gerlach VL, Shinkets RA, Zhong M;
XX Edinger SR, Ellerman K;
XX WPI; 2003-239445/23.
XX N-PSDB; ADC39165.
XX
XX New NOVX polypeptides and polynucleotides, useful in gene therapy,
XX particularly for treating or preventing a syndrome associated with a
XX human disease e.g. diabetes, obesity, cancer, Alzheimer's disease,
XX hypertension or hemophilia.
XX
XX Claim 1; SEQ ID NO 108; 748pp; English.
PS

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:11:52 ; Search time 6.86906 Seconds
(without alignments)
2152.832 Million cell updates/sec

Title: US-10-006-011A-9

Perfect score: 1566

Sequence: 1 CERQPCQHGATCMGAGEYEF.....QPLDLQHQAGANTRPCPS 284

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1566	100.0	4391	1	PGBM_HUMAN
2	1370.5	87.5	3707	1	PGBM_MOUSE
3	349	22.3	3375	1	UN52_CABEL
4	333.5	21.3	1959	1	AGRI_RAT
5	314.5	20.1	1328	1	AGRI_DISOM
6	301.5	19.3	1955	1	AGRI_CHICK
7	269	17.2	5147	1	FAT_DROME
8	253.5	16.2	1529	1	SLT2_HUMAN
9	242	15.5	4705	1	FAT2_DROME
10	225	14.4	3579	1	STAN_DROME
11	223.5	14.3	3110	1	LMX2_HUMAN
12	214	13.7	1578	1	NK3A_RAT
13	211.5	13.5	2923	1	CLR2_HUMAN
14	207.5	13.3	2144	1	CLR2_RAT
15	207.5	13.3	3106	1	LMX2_MOUSE
16	206.5	13.2	3312	1	CLR3_HUMAN
17	204.5	13.1	1712	1	NK2A_HUMAN
18	204.5	13.1	3313	1	CLR3_RAT
19	200	12.8	1541	1	NK3A_HUMAN
20	199	12.7	3321	1	NTC3_HUMAN
21	199	12.7	3014	1	CLR1_HUMAN
22	197.5	12.6	1530	1	NK1A_BOVIN
23	194.5	12.4	1715	1	NK2A_RAT
24	193	12.3	1363	1	NK1A_CHICK
25	193	12.3	1477	1	NK1A_HUMAN
26	193	12.3	1514	1	NK1A_RAT
27	191	12.2	3034	1	CLR1_MOUSE
28	190.5	12.2	618	1	DLI3_HUMAN
29	189.5	12.1	3301	1	CLR3_MOUSE
30	188.5	12.0	3075	1	LMX1_HUMAN
31	188	12.0	2215	1	CDN2_DROME
32	185.5	11.8	3672	1	LM12_CABEL
33	184	11.7	3097	1	CADN_DROME

34	183	11.7	1504	1	SLIT_DROME	P24014 drosophila
35	182.5	11.7	2003	1	NTC4_MOUSE	Q99466 homo sapien
36	182	11.6	2318	1	NTC3_MOUSE	Q61982 mus musculus
37	181.5	11.6	2531	1	NTC1_MOUSE	Q01705 mus musculus
38	181.5	11.6	2531	1	NTC1_RAT	Q07008 rattus norv
39	180.5	11.5	589	1	DLI3_RAT	O88671 rattus norv
40	180.5	11.5	592	1	DLI3_MOUSE	O88516 mus musculus
41	180.5	11.5	1247	1	JAG2_MOUSE	Q9QY65 mus musculus
42	180.5	11.5	1321	1	PGCN_HUMAN	O14594 homo sapien
43	180	11.5	1238	1	JAG2_HUMAN	Q9Y219 homo sapien
44	179.5	11.5	1202	1	JAG2_RAT	P97607 rattus norv
45	179.5	11.5	1242	1	JAG1_BRARE	Q90Y57 brachydanic

ALIGNMENTS

RESULT 1

PGBM_HUMAN	STANDARD;	PRT; 4391 AA.
ID	PGBM_HUMAN	
AC	P98160; Q16287; Q9H3V5;	
DT	01-OCT-1996 (Rel. 34, Created)	
DT	28-FEB-2003 (Rel. 41, Last sequence update)	
DT	10-OCT-2003 (Rel. 42, Last annotation update)	
DE	Basement membrane-specific heparan sulfate proteoglycan core	
DE	protein precursor (HSPG) (Perlecan) (PLC).	
GN	HSPG2.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=92112994; PubMed=1730768;	
RA	Kallunki P., Tryggvason K.;	
RT	"Human basement membrane heparan sulfate proteoglycan core protein: a	
RT	467-KD protein containing multiple domains resembling elements of the	
RT	low density lipoprotein receptor, laminin, neural cell adhesion	
RT	molecules, and epidermal growth factor."	
RL	J. Cell Biol. 116:559-571(1992).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RX	TISSUE=Colon, and Skin;	
RX	MEDLINE=92235084; PubMed=1569102;	
RA	Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;	
RT	"Primary structure of the human heparan sulfate proteoglycan from	
RT	basement membrane (HSPG2/perlecan): A chimeric molecule with multiple	
RT	domains homologous to the low density lipoprotein receptor, laminin,	
RT	neural cell adhesion molecules, and epidermal growth factor."	
RL	J. Biol. Chem. 267:8544-8557(1992).	
RN	[3]	
RP	SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SJS1 TYR-1532.	
RX	MEDLINE=20553141; PubMed=1101850;	
RA	Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barral D.,	
RA	Beighton P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,	
RA	Samson D., Urtizberea J.A., Lehmann-Horn F., Weissenbach J.,	
RA	Kentati F., Fontaine B.;	
RT	"Perlecan, the major proteoglycan of basement membranes, is altered in	
RT	patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia).";	
RL	Nat. Genet. 26:480-483(2000).	
RN	[4]	
RP	SEQUENCE OF 1016-1470 FROM N.A.	
RX	TISSUE=Colon;	
RX	MEDLINE=91365376; PubMed=1679749;	
RA	Dodge G.R., Kovacszy I., Chu M.L., Hassell J.R., McBride O.W.,	
RA	Xi H.F., Iozzo R.V.;	
RT	"Heparan sulfate proteoglycan of human colon: partial molecular	
RT	cloning, cellular expression, and mapping of the gene (HSPG2) to the	
RL	short arm of human chromosome 1.;"	
RL	Genomics 10:673-680(1991).	
RN	[5]	
RP	SEQUENCE OF 890-1396 FROM N.A.	
RX	TISSUE=Fibrosarcoma;	

CC RX MEDLINE=92120660; PubMed=1685141;
RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA Trygvaeson K.;
RT "Cloning of human heparan sulfate proteoglycan core protein,
RT assignment of the gene (HSPG2) to 1p36.1--p35 and identification of
RT a BamHI restriction fragment length polymorphism.";
RL Genomics 11:389-396(1991).
RN [6]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=94052171; PubMed=8234307;
RA Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RT "Structural characterization of the complete human perlecan gene and
RT its promoter.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
RN [7]
RP CARBOHYDRATE-LINKAGE SITE ASN-2121.
RX MEDLINE=22660472; PubMed=12754519;
RA Zhang H., Li X.-J., Martin D.B., Aebbersold R.;
RT "Identification and quantification of N-linked glycoproteins using
RT hydrazide chemistry, stable isotope labeling and mass spectrometry.";
RL Nat. Biotechnol. 21:660-666(2003).
CC -!- FUNCTION: This protein is an integral component of basement
CC membranes. It is responsible for the fixed negative electrostatic
CC charge and is involved in the charge-selective ultrafiltration
CC properties. It serves as an attachment substrate for cells.
CC -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC dimers or stellate structures. It interacts with other basement
CC membrane components such as laminin, prolargin and collagen type
CC IV.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Found in the basement membranes.
CC -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -!- DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel
CC syndrome (JSJ) [MIM:255800]; a rare autosomal recessive disorder
CC characterized by permanent myotonia (prolonged failure of muscle
CC relaxation) and skeletal dysplasia, resulting in reduced stature,
CC kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.
CC -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC -!- SIMILARITY: Contains 11 laminin EGF-like domains.
CC -!- SIMILARITY: Contains 3 laminin IV domains.
CC -!- SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 3 laminin G-like domains.
CC -!- SIMILARITY: Contains 4 EGF-like domains.
CC -!- SIMILARITY: Contains 1 SEA domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X62515; CAA44373.1; -;
CC EMBL; M95289; AA52700.1; -;
CC EMBL; AL45795; CAC18534.1; -;
CC EMBL; M64283; AA52699.1; -;
CC EMBL; S76436; AAB21121.2; -;
CC EMBL; L22078; -; NOT ANNOTATED_CDS.
CC PIR; P00740; LEDN.
CC SIENA-2DPAGE; P98160; -;
CC Genew; HGNC:5273; HSPG2.
CC MIM; 142461; -;
CC MIM; 255800; -;
CC InterPro; IPR008985; ConA like_lec_gl.
CC InterPro; IPR000742; EGF 2.
CC InterPro; IPR006209; EGF like.
CC InterPro; IPR006210; IEGF.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003599; Ig.
CC InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003596; Ig_v.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR002172; LDL receptor_A.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00047; Ig; 22.
DR Pfam; PF00052; laminin_B; 3.
DR Pfam; PF00053; laminin_EGF; 7.
DR Pfam; PF00054; laminin_G; 3.
DR Pfam; PF00057; ldl_recept_a; 4.
DR Pfam; PF01390; SEA; 1.
DR PRINTS; PR00261; LDLRECEPTOR.
DR ProDom; PD003031; Laminin_B; 3.
DR SMART; SM00181; EGF; 15.
DR SMART; SM00180; EGF_Lam; 12.
DR SMART; SM00409; IG; 22.
DR SMART; SM00408; IGC2; 21.
DR SMART; SM00406; IGV; 7.
DR SMART; SM00281; LamB; 3.
DR SMART; SM00282; LamG; 3.
DR SMART; SM00192; LDLa; 4.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF 1; 9.
DR PROSITE; PS01186; EGF 2; 6.
DR PROSITE; PS00026; EGF 3; 4.
DR PROSITE; PS00835; IG LIKE; 22.
DR PROSITE; PS50025; LAM G DOMAIN; 3.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR PROSITE; PS01209; LDLRA_1; 4.
DR PROSITE; PS00068; LDLRA_2; 4.
DR PROSITE; PS50024; SEA; 1.
KW Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW Extracellular matrix; EGF-like domain; Disease mutation.
FT SIGNAL 1 21
FT CHAIN 22 4391
FT BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT SULFATE PROTEOGLYCAN CORE PROTEIN.
FT SEA.
FT DOMAIN 80 194
FT LDL-RECEPTOR CLASS A 1.
FT DOMAIN 198 235
FT LDL-RECEPTOR CLASS A 2.
FT DOMAIN 284 320
FT LDL-RECEPTOR CLASS A 3.
FT DOMAIN 324 360
FT LDL-RECEPTOR CLASS A 4.
FT DOMAIN 367 404
FT IG-LIKE C2-TYPE 1
FT DOMAIN 405 504
FT LAMININ EGF-LIKE 1 (N-TERMINAL).
FT DOMAIN 521 530
FT LAMININ DOMAIN IV 1 (DOMAIN III A).
FT DOMAIN 531 730
FT LAMININ EGF-LIKE 1 (C-TERMINAL).
FT DOMAIN 731 763
FT LAMININ EGF-LIKE 2.
FT DOMAIN 764 813
FT LAMININ EGF-LIKE 3.
FT DOMAIN 814 871
FT LAMININ EGF-LIKE 4. (INCOMPLETE).
FT DOMAIN 879 923
FT LAMININ EGF-LIKE 5 (N-TERMINAL).
FT DOMAIN 924 933
FT LAMININ DOMAIN IV 2 (DOMAIN III B).
FT DOMAIN 934 1125
FT LAMININ EGF-LIKE 6.
FT DOMAIN 1126 1158
FT LAMININ EGF-LIKE 7.
FT DOMAIN 1159 1208
FT LAMININ EGF-LIKE 8.
FT DOMAIN 1209 1265
FT LAMININ EGF-LIKE 9 (N-TERMINAL).
FT DOMAIN 1275 1324
FT LAMININ EGF-LIKE 10.
FT DOMAIN 1325 1334
FT LAMININ DOMAIN IV 3 (DOMAIN III C).
FT DOMAIN 1335 1529
FT LAMININ EGF-LIKE 11.
FT DOMAIN 1530 1562
FT LAMININ EGF-LIKE 12.
FT DOMAIN 1563 1612
FT IG-LIKE C2-TYPE 2.
FT DOMAIN 1613 1670
FT IG-LIKE C2-TYPE 3.
FT DOMAIN 1677 1771
FT IG-LIKE C2-TYPE 4.
FT DOMAIN 1772 1865
FT IG-LIKE C2-TYPE 5.
FT DOMAIN 1866 1955
FT IG-LIKE C2-TYPE 6.
FT DOMAIN 1956 2051
FT IG-LIKE C2-TYPE 7.
FT DOMAIN 2052 2244
FT IG-LIKE C2-TYPE 8.
FT DOMAIN 2245 2340
FT IG-LIKE C2-TYPE 9.
FT DOMAIN 2341 2436
FT IG-LIKE C2-TYPE 10.
FT DOMAIN 2437 2533
FT IG-LIKE C2-TYPE 11.
FT DOMAIN 2534 2629
FT IG-LIKE C2-TYPE 12.


```

FT DOMAIN 879 923 LAMININ EGF-LIKE 4 (INCOMPLETE).
FT DOMAIN 924 933 LAMININ EGF-LIKE 5 (N-TERMINAL).
FT DOMAIN 934 1125 LAMININ DOMAIN IV 2 (DOMAIN III B).
FT DOMAIN 1126 1158 LAMININ EGF-LIKE 5 (C-TERMINAL).
FT DOMAIN 1159 1208 LAMININ EGF-LIKE 6.
FT DOMAIN 1209 1265 LAMININ EGF-LIKE 7.
FT DOMAIN 1275 1324 LAMININ EGF-LIKE 8.
FT DOMAIN 1325 1334 LAMININ EGF-LIKE 9 (N-TERMINAL).
FT DOMAIN 1335 1529 LAMININ DOMAIN IV 3 (DOMAIN III C).
FT DOMAIN 1530 1562 LAMININ EGF-LIKE 9 (C-TERMINAL).
FT DOMAIN 1563 1610 LAMININ EGF-LIKE 10.
FT DOMAIN 1611 1670 LAMININ EGF-LIKE 11.
FT DOMAIN 1671 1771 IG-LIKE C2-TYPE 2.
FT DOMAIN 1772 1865 IG-LIKE C2-TYPE 3.
FT DOMAIN 1866 1954 IG-LIKE C2-TYPE 4.
FT DOMAIN 1955 2049 IG-LIKE C2-TYPE 5.
FT DOMAIN 2050 2148 IG-LIKE C2-TYPE 6.
FT DOMAIN 2149 2243 IG-LIKE C2-TYPE 7.
FT DOMAIN 2244 2343 IG-LIKE C2-TYPE 8.
FT DOMAIN 2344 2436 IG-LIKE C2-TYPE 9.
FT DOMAIN 2437 2532 IG-LIKE C2-TYPE 10.
FT DOMAIN 2533 2619 IG-LIKE C2-TYPE 11.
FT DOMAIN 2620 2720 IG-LIKE C2-TYPE 12.
FT DOMAIN 2721 2809 IG-LIKE C2-TYPE 13.
FT DOMAIN 2810 2895 IG-LIKE C2-TYPE 14.
FT DOMAIN 2896 2980 IG-LIKE C2-TYPE 15.
FT DOMAIN 2981 3162 LAMININ G-LIKE 1.
FT DOMAIN 3163 3241 EGF-LIKE.
FT DOMAIN 3242 3425 LAMININ G-LIKE 2.
FT DOMAIN 3426 3705 LAMININ G-LIKE 3.
FT SITE 65 67 HEPARAN SULFATE (POTENTIAL).
FT SITE 71 73 HEPARAN SULFATE (POTENTIAL).
FT SITE 76 78 HEPARAN SULFATE (POTENTIAL).
FT SITE 3615 3617 MEDIATES MOTOR NEURON ATTACHMENT (POTENTIAL).

FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 206 225 BY SIMILARITY.
FT DISULFID 229 234 BY SIMILARITY.
FT DISULFID 285 297 BY SIMILARITY.
FT DISULFID 292 310 BY SIMILARITY.
FT DISULFID 304 319 BY SIMILARITY.
FT DISULFID 325 337 BY SIMILARITY.
FT DISULFID 332 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 368 381 BY SIMILARITY.
FT DISULFID 375 394 BY SIMILARITY.
FT DISULFID 388 403 BY SIMILARITY.
FT DISULFID 428 479 BY SIMILARITY.
FT DISULFID 764 773 BY SIMILARITY.
FT DISULFID 766 780 BY SIMILARITY.
FT DISULFID 783 792 BY SIMILARITY.
FT DISULFID 795 811 BY SIMILARITY.
FT DISULFID 844 829 BY SIMILARITY.
FT DISULFID 846 839 BY SIMILARITY.
FT DISULFID 842 851 BY SIMILARITY.
FT DISULFID 854 869 BY SIMILARITY.
FT DISULFID 1159 1168 BY SIMILARITY.
FT DISULFID 1161 1175 BY SIMILARITY.
FT DISULFID 1178 1187 BY SIMILARITY.
FT DISULFID 1190 1206 BY SIMILARITY.
FT DISULFID 1209 1224 BY SIMILARITY.
FT DISULFID 1211 1234 BY SIMILARITY.
FT DISULFID 1237 1246 BY SIMILARITY.
FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1275 1287 BY SIMILARITY.
FT DISULFID 1277 1293 BY SIMILARITY.
FT DISULFID 1295 1304 BY SIMILARITY.
FT DISULFID 1307 1322 BY SIMILARITY.
FT DISULFID 1563 1572 BY SIMILARITY.
FT DISULFID 1565 1579 BY SIMILARITY.
FT DISULFID 1582 1591 BY SIMILARITY.
FT DISULFID 1594 1610 BY SIMILARITY.
FT DISULFID 1613 1628 BY SIMILARITY.

Query Match 87.5%; Score 1370.5; DB 1; Length 3707;
Best Local Similarity 86.3%; Pred. No. 1.4e-89;
Matches 245; Conservative 20; Mismatches 18; Indels 1; Gaps 1;

QY 1 CERPCQCHGATCMPAGEYEFQCLCRDGRYKGLCEHEENPCQLREPCCLHGTGCTGTRCLCL 60
DB 3425 CERPCRNATCMPAGEYEFQCLCQDGFGRDLCHEENPCQLHEPCINGTCRGARCLCL 3484

QY 61 PGFSGPRCCQGGCHGIAESDWHLEGSGGNDARPGQYGFHDDGFLAPPGHVSRLPEVP 120
DB 3485 PGFSGPRCCQGGAGYGVVESDWHPEGGSGNDAPGQYGFYDNGFLGCPGNSFSRLPEVP 3544

QY 121 ETIELEVRSTASGLLLWQGVGEAGQKDPISLGLQDHLVFRYQLSGSEARLYSEDP 180
DB 3545 ETIEFEVRTSTADGLLLWQGV-VREASRSKDPISLGLQDHLVFRYQLSGSEARLYVSGDP 3603

QY 181 INDGEWHRTALREGRGRGSIQVDGELVSGRPGPNVAVNAKSGSVYIGGAPDVATITGGR 240
DB 3604 INDGEWHRTALREGRGRGSIQVDGELVSGRPGPNVAVNTKDIYIGGAPDVATITRCK 3663

QY 241 FSSGITGVKNVLVHSARGAPPQPPQLDQHRQAQAGANTRPCPS 284
DB 3664 FSSGITGVKNVLVHSARGAPPQPPQLDQHRQAQAGANTRPCPS 3707

RESULT 3
UN52_CABEL STANDARD; PRT: 3375 AA.
ID UN52_CABEL 018263; Q9XTD2; Q9XT15;
AC Q06561; O18263; Q9XTD2; Q9XT15;
DT 01-JUN-1994 (Rel. 29, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane proteoglycan precursor (Perlecan homolog)
DE (Uncoordinated protein 52).
GN UNC-52 OR ZC101.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM A), AND FUNCTION.
RX MEDLINE=9339574; PubMed=8393416;
RA Rogalski T.M., Williams B.D., Mullen G.P., Moerman D.G.;
RT "Products of the unc-52 gene in Caenorhabditis elegans are homologous to the core protein of the mammalian basement membrane heparan sulfate proteoglycan."
RN Genes Dev. 7:1471-1484(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Percy C.M., Baynes C.;
RN Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RA Durbin R.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -/- FUNCTION: Probable role in myofilament assembly and/or attachment of the myofilament lattice to the cell membrane. May be an extracellular anchor for integrin receptors in muscle.
CC -/- SUBCELLULAR LOCATION: Extracellular matrix.
CC -/- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=;

```

CC IsoId=Q06561-1; Sequence=VSP_007195, VSP_007196;
 CC Note=No experimental confirmation available;
 CC Name=a;
 CC IsoId=Q06561-2; Sequence=VSP_007195, VSP_007196;
 CC Name=b;
 CC IsoId=Q06561-3; Sequence=VSP_007191, VSP_007192;
 CC Note=No experimental confirmation available;
 CC Name=c;
 CC IsoId=Q06561-4; Sequence=VSP_007193, VSP_007194, VSP_007195,
 CC VSP_007196;
 CC Note=No experimental confirmation available;
 CC -!- TISSUE SPECIFICITY: Found in the basement membrane of all
 CC contractile tissues. It is concentrated over muscle dense bodies
 CC and M-lines which are associated with beta-integrin.
 CC -!- DEVELOPMENTAL STAGE: Synthesized early in embryogenesis.
 CC -!- SIMILARITY: Contains 3 LDL-receptor class A domains.
 CC -!- SIMILARITY: Contains 16 immunoglobulin-like C2-type domains.
 CC -!- SIMILARITY: Contains 7 laminin EGF-like domains.
 CC -!- SIMILARITY: Contains 3 laminin G-like domains.
 CC -!- SIMILARITY: Contains 2 laminin IV domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L13458; AAA28156.1; -;
 CC EMBL; Z93375; CAB07567.1; -;
 CC EMBL; Z93395; CAB07567.1; JOINED.
 CC EMBL; Z93375; CAB07568.1; -;
 CC EMBL; Z93395; CAB07568.1; JOINED.
 CC EMBL; Z93375; CAB07569.1; -;
 CC EMBL; Z93395; CAB07569.1; JOINED.
 CC EMBL; Z93395; CAB07704.1; -;
 CC EMBL; Z93395; CAB07706.1; -;
 CC EMBL; Z93375; CAB07706.1; JOINED.
 CC EMBL; Z93395; CAB07707.1; -;
 CC EMBL; Z93375; CAB07707.1; JOINED.
 CC EMBL; Z93395; CAB07708.1; -;
 CC EMBL; Z93375; CAB07708.1; JOINED.
 CC HSP; P01130; ILDR.
 CC WormPep; ZC101.2a; CE15028.
 CC WormPep; ZC101.2b; CE15030.
 CC WormPep; ZC101.2c; CE15034.
 CC WormPep; ZC101.2e; CE18424.
 CC GO; GO:0005578; C:extracellular matrix; IEP.
 CC GO; GO:0030239; P:myofibril assembly; IEP.
 CC InterPro; IPR008985; ConA_like_1ec_gi.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR006209; EGF_Like.
 CC InterPro; IPR006210; IEGF.
 CC InterPro; IPR007110; Ig-Like.
 CC InterPro; IPR003599; Ig.
 CC InterPro; IPR003598; Ig_c2.
 CC InterPro; IPR000034; Laminin_B.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR001791; Laminin_G.
 CC InterPro; IPR002172; LDL_receptor_A.
 CC Pfam; PF00047; ig; 16
 CC Pfam; PF00052; laminin_B; 2.
 CC Pfam; PF00053; laminin_EGF; 5.
 CC Pfam; PF00057; ldl_recept_a; 3.
 CC PRINTS; PR00261; LDLRECEPTOR.
 CC ProDom; PD003031; Laminin_B; 2.
 CC SMART; SM00181; EGF; 6.
 CC SMART; SM00179; EGF_Ca; 2.
 CC SMART; SM00180; EGF_Lam; 6.
 CC SMART; SM00409; IG; 17.
 CC SMART; SM00408; IGC2; 17.
 CC SMART; SM00281; LamB; 1.

DR SMART; SM00282; LamG; 3.
 DR SMART; SM00192; LDLa; 3.
 DR PROSITE; PS00022; EGF_1; 7.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS00026; EGF_3; 3.
 DR PROSITE; PS00835; IG_Like; 17.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
 DR PROSITE; PS01209; LDLa_1; 3.
 DR PROSITE; PS00681; LDLa_2; 3.
 KW Proteoglycan; Glycoprotein; Signal; Immunoglobulin domain; Repeat;
 KW Basement membrane; Extracellular matrix; Alternative splicing;
 KW Laminin EGF-like domain.
 FT SIGNAL; 1 22
 FT CHAIN; 23 3375
 FT DOMAIN; 45 130
 FT DOMAIN; 148 184
 FT DOMAIN; 189 225
 FT DOMAIN; 232 269
 FT DOMAIN; 271 355
 FT DOMAIN; 384 431
 FT DOMAIN; 432 441
 FT DOMAIN; 442 633
 FT DOMAIN; 634 666
 FT DOMAIN; 674 720
 FT DOMAIN; 721 730
 FT DOMAIN; 731 921
 FT DOMAIN; 922 954
 FT DOMAIN; 955 1004
 FT DOMAIN; 1011 1060
 FT DOMAIN; 1061 1111
 FT DOMAIN; 1126 1222
 FT DOMAIN; 1226 1311
 FT DOMAIN; 1319 1401
 FT DOMAIN; 1410 1499
 FT DOMAIN; 1503 1585
 FT DOMAIN; 1589 1680
 FT DOMAIN; 1690 1785
 FT DOMAIN; 1793 1878
 FT DOMAIN; 1886 1970
 FT DOMAIN; 1973 2069
 FT DOMAIN; 2073 2173
 FT DOMAIN; 2173 2260
 FT DOMAIN; 2263 2343
 FT DOMAIN; 2349 2435
 FT DOMAIN; 2446 2530
 FT DOMAIN; 2532 2713
 FT DOMAIN; 2793 2960
 FT DOMAIN; 2961 3093
 FT DOMAIN; 2972 3066
 FT DOMAIN; 3180 3359
 FT DISULFID; 66 114
 FT DISULFID; 149 161
 FT DISULFID; 156 174
 FT DISULFID; 168 183
 FT DISULFID; 190 202
 FT DISULFID; 197 215
 FT DISULFID; 209 224
 FT DISULFID; 233 246
 FT DISULFID; 240 259
 FT DISULFID; 253 268
 FT DISULFID; 253 268
 FT DISULFID; 955 964
 FT DISULFID; 957 971
 FT DISULFID; 974 983
 FT DISULFID; 986 1002
 FT DISULFID; 1011 1021
 FT DISULFID; 1013 1027
 FT DISULFID; 1030 1039
 FT DISULFID; 1042 1058
 FT DISULFID; 1061 1069
 FT DISULFID; 1063 1079
 FT DISULFID; 1082 1091
 FT DISULFID; 1094 1109

FT DISULFID 1152 1200 BY SIMILARITY.
FT DISULFID 1384 BY SIMILARITY.
FT DISULFID 1435 1481 BY SIMILARITY.
FT DISULFID 1527 1573 BY SIMILARITY.
FT DISULFID 1618 1663 BY SIMILARITY.
FT DISULFID 1719 1767 BY SIMILARITY.
FT DISULFID 1814 1861 BY SIMILARITY.
FT DISULFID 1907 1954 BY SIMILARITY.
FT DISULFID 1998 2053 BY SIMILARITY.
FT DISULFID 2099 2147 BY SIMILARITY.
FT DISULFID 2195 2242 BY SIMILARITY.
FT DISULFID 2284 2329 BY SIMILARITY.
FT DISULFID 2374 2420 BY SIMILARITY.
FT DISULFID 2467 2514 BY SIMILARITY.
FT CARBOHYD 1422 1422 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2476 2476 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2950 2950 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3143 3143 N-LINKED (GLCNAC. .) (POTENTIAL).

Query Match 22.3%; Score 349; DB 1; Length 3375;
Best Local Similarity 31.2%; Pred. No. 6.5e-17;
Matches 81; Conservative 41; Mismatches 102; Indels 36; Gaps 6;

QY 30 GDLCHHEB--NPCQUREPLHGTCQGR-----CLLPFGSPRCQCGSHGIAESDWH 82
Db 3129 GDVSTQEPNNIC-ANSTCGMNGQCPVPMTHYTCECKLYYDGTCSLFK----- 3177
QY 83 LEGSGNDAPGGYGFYHDDGFLAFFGHVFSRSLPEVPEETLELVRTSTASGLLLWQVE 142
Db 3178 -----PIEHAARFGDAFIELSSDEFPHLISEKDEIVAPKFKTEQQNVLLWQV-Q 3227
QY 143 VGEAQGQKDFSLGLODGHVFRYVGLSGEARLVSEDPDNGEWHRYVALREGSRGIQV 202
Db 3228 RPTVQOMEDYISVGIVNGHLFHSVELGGAALHSEERVDGKHSVRFERKSGREGQMRI 3287
QY 203 DGEELVSRSPQPVANNAKSVWIGGAPDVATLTGGRFFSGITGCYKVLVLSARPAP 262
Db 3288 DNYREVDGRSGTILAMVNDGNIFVGGVDFDKATGGLFSNNFGCCADVELNGVK---- 3343
QY 263 PPQPLDLQHRAQAGANTRPC 282
Db 3344 ----LDLMATAIDGKNVRPC 3359

RESULT 4
AGRI_RAT STANDARD; PRT; 1959 AA.
AC P25304; Q63034;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Agrin precursor.
GN AGRN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 1-1779 AND 1799-1959 FROM N.A.
RC TISSUE=embryonic spinal cord;
RX MEDLINE=9122570; PubMed=1851019;
RA Rupp F., Payan D.G., Magill-Solc C., Cowan D.M., Scheller R.H.;
RT "Structure and expression of a rat agrin.";
RL Neuron 6:811-823(1991).
RN [2]
RP SEQUENCE OF 1777-1801 FROM N.A.
RX MEDLINE=92407628; PubMed=1326508;
RA Rupp F., Oezcelik T., Linial M.; Peterson K., Francke U., Scheller R.;
RT "Structure and chromosomal localization of the mammalian agrin gene.";
RL J. Neurosci. 12:3535-3544(1992).
CC -!- FUNCTION: Component of the basal lamina that causes the
aggregation of acetylcholine receptors and acetylcholine-esterase
on the surface of muscle fibers of the neuromuscular junction.

CC -!- SUBUNIT: Binds to laminin.
CC -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
junction.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Comment=Additional isoforms seem to exist. Isoforms differ in
their acetylcholine receptor clustering activity;
CC Name=1;
CC IsoId=P25304-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P25304-2; Sequence=VSP_001365;
CC Name=3;
CC IsoId=P25304-3; Sequence=VSP_001366;
CC Name=4;
CC IsoId=P25304-4; Sequence=VSP_001367;
CC Name=5;
CC IsoId=P25304-5; Sequence=VSP_001368;
CC -!- TISSUE SPECIFICITY: Embryonic nervous system and muscle.
CC -!- DEVELOPMENTAL STAGE: More abundant early in development.
CC -!- PTM: Contains heparan sulfate chains as well as N-linked and O-
linked oligosaccharides (By similarity).
CC -!- SIMILARITY: Contains 9 Kazal-like domains.
CC -!- SIMILARITY: Contains 2 laminin EGF-like domains.
CC -!- SIMILARITY: Contains 4 EGF-like domains.
CC -!- SIMILARITY: Contains 1 SEA domain.
CC -!- SIMILARITY: Contains 3 laminin G-like domains.
CC -!- CAUTION: It is uncertain whether Met-1, Met-18 or Met-24 is the
initiator.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M64780; AAA40703.1; -;
CC EMBL; M64780; AAA40702.1; ALT_INIT.
CC EMBL; S44194; AAB23326.1; -;
CC PIR; JH0399; AGRT.
CC HSSP; P00740; 1EDM.
CC InterPro; IPR008985; ConA_like_lec_gl.
CC InterPro; IPR006209; EGF_like.
CC InterPro; IPR003645; FOLN.
CC InterPro; IPR002350; Kazal.
CC InterPro; IPR002049; Laminin_EGF.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR000082; SEA_domain.
CC Pfam; PF00008; EGF; 4.
CC Pfam; PF00050; kazal; 9.
CC Pfam; PF00053; laminin_EGF; 2.
CC Pfam; PF00054; laminin_G; 3.
CC Pfam; PF01390; SEA; 1.
CC PRINTS; PRC0011; EGFLAMININ.
CC SMART; SM00180; EGF_Lam; 2.
CC SMART; SM00274; FOLN; 8.
CC SMART; SM00280; KAZAL; 9.
CC SMART; SM0282; lamG; 3.
CC SMART; SM0200; SEA; 1.
CC PROSITE; PS00022; EGF_1; 6.
CC PROSITE; PS01186; EGF_2; 1.
CC PROSITE; PS00026; EGF_3; 4.
CC PROSITE; PS00025; LAM_G_DOMAIN; 3.
CC PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
CC PROSITE; PS00024; SEA; 1.
KW Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
KW Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
FT SIGNAL 1 29
FT CHAIN 30 1959
FT DOMAIN 65 137 KAZAL-LIKE 1.
FT DOMAIN 141 212 KAZAL-LIKE 2.
FT DOMAIN 213 284 KAZAL-LIKE 3.

FT DOMAIN 287 356 KAZAL-LIKE 4.
 FT DOMAIN 361 429 KAZAL-LIKE 5.
 FT DOMAIN 430 494 KAZAL-LIKE 6.
 FT DOMAIN 495 559 KAZAL-LIKE 7.
 FT DOMAIN 563 645 KAZAL-LIKE 8.
 FT DOMAIN 688 741 LAMININ EGF-LIKE 1.
 FT DOMAIN 742 788 LAMININ EGF-LIKE 2.
 FT DOMAIN 794 864 KAZAL-LIKE 9.
 FT DOMAIN 1023 1145 SEA.
 FT DOMAIN 1220 1258 EGF-LIKE 1.
 FT DOMAIN 1263 1439 LAMININ G-LIKE 1.
 FT DOMAIN 1440 1477 EGF-LIKE 2.
 FT DOMAIN 1479 1516 EGF-LIKE 3.
 FT DOMAIN 1526 1708 LAMININ G-LIKE 2.
 FT DOMAIN 1709 1748 EGF-LIKE 4.
 FT DOMAIN 1784 1956 LAMININ G-LIKE 3.
 FT DOMAIN 869 992 SER/THR-RICH.
 FT DOMAIN 1147 1215 SER/THR-RICH.
 FT DISULFID 97 116 POTENTIAL.
 FT DISULFID 105 137 POTENTIAL.
 FT DISULFID 171 191 POTENTIAL.
 FT DISULFID 180 212 POTENTIAL.
 FT DISULFID 244 263 POTENTIAL.
 FT DISULFID 252 284 POTENTIAL.
 FT DISULFID 316 335 POTENTIAL.
 FT DISULFID 324 356 POTENTIAL.
 FT DISULFID 389 408 POTENTIAL.
 FT DISULFID 397 429 POTENTIAL.
 FT DISULFID 454 473 POTENTIAL.
 FT DISULFID 462 494 POTENTIAL.
 FT DISULFID 518 538 POTENTIAL.
 FT DISULFID 527 559 POTENTIAL.
 FT DISULFID 604 624 POTENTIAL.
 FT DISULFID 613 645 POTENTIAL.
 FT DISULFID 688 700 BY SIMILARITY.
 FT DISULFID 690 707 BY SIMILARITY.
 FT DISULFID 709 718 BY SIMILARITY.
 FT DISULFID 721 739 BY SIMILARITY.
 FT DISULFID 742 754 BY SIMILARITY.
 FT DISULFID 744 761 BY SIMILARITY.
 FT DISULFID 763 772 BY SIMILARITY.
 FT DISULFID 775 786 BY SIMILARITY.
 FT DISULFID 823 843 POTENTIAL.
 FT DISULFID 832 864 POTENTIAL.
 FT DISULFID 1224 1235 BY SIMILARITY.
 FT DISULFID 1229 1246 BY SIMILARITY.
 FT DISULFID 1248 1257 BY SIMILARITY.
 FT DISULFID 1444 1455 POTENTIAL.
 FT DISULFID 1449 1465 POTENTIAL.
 FT DISULFID 1467 1476 POTENTIAL.
 FT DISULFID 1483 1494 BY SIMILARITY.
 FT DISULFID 1488 1504 BY SIMILARITY.
 FT DISULFID 1506 1515 BY SIMILARITY.
 FT DISULFID 1713 1727 BY SIMILARITY.
 FT DISULFID 1721 1736 BY SIMILARITY.
 FT DISULFID 1738 1747 BY SIMILARITY.
 FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 672 672 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 957 957 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 1144 1152 Missing (in isoform 2).
 FT VARSPLIC 1780 1798 Missing (in isoform 3).
 FT VARSPLIC 1788 1798 Missing (in isoform 4).
 FT VARSPLIC 1780 1787 Missing (in isoform 5).
 FT VARSPLIC 314 314 V -> VTCD (IN A VARIANT).
 FT SEQUENCE 1959 AA; 208645 MW; 7FEFDFAFF89CC31 CRC64;
 Query Match 21.3%; Score 333.5; DB 1; Length 1959;
 Best Local Similarity 36.0%; Pred. No. 4.5e-16;

Matches 94; Conservative 28; Mismatches 108; Indels 31; Gaps 10;
 QY 1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGLCEHENPCQLRFPCLHGGTCQ-----GT 55
 DB 1444 CLPNECHGGLCALQALEAGMFLCQCQPPGRFGPTCADEKSPQ-Q-PNFCGAAFCRVLSGGA 1502
 QY 56 RCLCLPGSPGRCQCGSGHGAESDWHLEGSGGNDAPQCYGAYFHDDGFLAFPG-HVFSR 114
 DB 1503 KCECPGRSGTFCOT-----VLETAGSR-----PFLADFNFSYLEKGLHTFER 1547
 QY 115 SIPEVPETIELEVRTSTASGLLLWQGVVEGAGQCKDFISLQDQGHVFRYQLGSGEAR 174
 DB 1548 DLGE-KMALEWVFLARGPSGLLYNGQKTD--GKG-DFVSLAHNRHLEFCYDLGKGA 1603
 QY 175 LVSEDPINDGEWHRVTALREGRSGIQVDGEELVSGRSPG-----PNVAVNAKSGVYIGGA 230
 DB 1604 IRSKEPIALGTWVRVFLERNRKGALQVGDGFRVLGSPKSRKVPHTMLNLKEPLYIGGA 1663
 QY 231 PDVATLT-GRFPSSGITGVK 250
 DB 1664 PDFSKLARGAAVSSGFGSVIQ 1684
 RESULT 5
 AGRI DISOM
 ID AGRI DISOM STANDARD; PRT; 1328 AA.
 AC Q90404;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Agrin (Fragment).
 GN AGRN.
 OS Discoppyge ommata (Electric ray).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Squala; Hypnosqualea; Pristiogaster; Batoidae;
 OC Torpediniformes; Narcinidae; Narcinidae; Discoppyge.
 OX NCBI_taxid=7785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smith M.A., Magill-Solc C., Rupp F., Yao Y.-M.M., Schilling J.W.,
 RA Snow P., McMahon U.J.;
 RT "Isolation and characterization of a cDNA that encodes an agrin
 homolog in the marine ray.";
 RL Mol. Cell. Neurosci. 3:406-417(1992).
 CC -!- FUNCTION: Component of the basal lamina that causes the
 CC aggregation of acetylcholine receptors and acetylcholine-esterase
 CC on the surface of muscle fibers of the neuromuscular junction (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
 CC junction (By similarity).
 CC -!- SIMILARITY: Contains at least 2 laminin EGF-like domains.
 CC -!- SIMILARITY: Contains at least 3 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 SEA domain.
 CC -!- SIMILARITY: Contains 3 laminin G-like domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; L01423; AAA49224.1; --
 CC PIR; T43060; T43060.
 CC HSSP; P00740; IEDM.
 CC InterPro; IPR00152; Asx_hydroxyl_S.
 CC InterPro; IPR008985; ConA_like_lec_gl.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR006209; EGF_Like.
 CC InterPro; IPR003645; PolN.
 CC InterPro; IPR002350; kazal.
 CC InterPro; IPR002049; laminin_EGF.

```
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00050; Kazal; 2.
DR Pfam; PF00053; laminin_EGF; 2.
DR Pfam; PF00054; laminin_G; 3.
DR Pfam; PF01390; SEA; 1.
DR PRINTS; PR00011; EGF_LAMININ.
DR SMART; SMO0180; EGF_Lam; 2.
DR SMART; SMO0274; FOLN; 2.
DR SMART; SMO0280; KAZAL; 2.
DR SMART; SMO0282; LamG; 3.
DR SMART; SMO0200; SEA; 1.
DR PROSITE; PS00010; ASK_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 5.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00026; EGF_3; 4.
DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR PROSITE; PS00024; SEA; 1.
DR Glycoprotein; EGF-like domain; Repeat; Laminin EGF-like domain.
FT NON_TER 1
FT DOMAIN 79 132 LAMININ EGF-LIKE 1.
FT DOMAIN 133 179 LAMININ EGF-LIKE 2.
FT DOMAIN 411 533 SEA.
FT DOMAIN 608 644 EGF-LIKE 1.
FT DOMAIN 649 825 LAMININ G-LIKE 1.
FT DOMAIN 865 902 EGF-LIKE 2.
FT DOMAIN 914 1096 LAMININ G-LIKE 2.
FT DOMAIN 1097 1135 EGF-LIKE 3.
FT DOMAIN 1146 1324 LAMININ G-LIKE 3.
FT DISULFID 79 91 BY SIMILARITY.
FT DISULFID 81 98 BY SIMILARITY.
FT DISULFID 100 109 BY SIMILARITY.
FT DISULFID 112 130 BY SIMILARITY.
FT DISULFID 133 145 BY SIMILARITY.
FT DISULFID 135 152 BY SIMILARITY.
FT DISULFID 154 163 BY SIMILARITY.
FT DISULFID 166 177 BY SIMILARITY.
FT DISULFID 612 623 BY SIMILARITY.
FT DISULFID 617 632 BY SIMILARITY.
FT DISULFID 634 643 BY SIMILARITY.
FT DISULFID 859 880 BY SIMILARITY.
FT DISULFID 874 890 BY SIMILARITY.
FT DISULFID 892 901 BY SIMILARITY.
FT DISULFID 1101 1114 BY SIMILARITY.
FT DISULFID 1108 1123 BY SIMILARITY.
FT DISULFID 1125 1134 BY SIMILARITY.
FT CARBOHYD 63 63 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 279 279 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 369 369 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1328 AA; 144018 MW; 79DB1C1AF2A71C18 CRC64;

Query Match
Best Local Similarity 20.1%; Score 314.5; DB 1; Length 1328;
Matches 85; Conservative 44; Mismatches 105; Indels 37; Gaps 12;

QY 1 CERQPCQHGATCMPAGEYFQCLCRDGFKGLCEHEENPCQLRPFCLHGTCQ-----GT 55
DB 830 CSPNPKNRKCHKMKEAEMFHCSEYGFSGPTCDKHNPCD-PNPFQSANCWVLPPEGGS 888
QY 56 RCLCLPGSPGRCQSGHGTAEBSWHLEGGSGNDAPQGYAYFHDGFLAPPG-HVFSR 114
DB 889 KCECPMGRGEELUCER-----VSEAE-----QQQKAFIPFNGLSYLNMGIHTF-- 933
QY 115 SLFVEPTIELEV--RTSTAGLLWQGVGEAGCGKDFISLGLQDGLHVFYRLQSGGE 172
DB 934 -VSDLLQKLSMEVIFLAKDPNGMIFYNGQKTD--GRG-DFVSLNLRDGLYLFKYDLKGKA 989
QY 173 ARLVSEDPIDNGEWRVYALRGGRGSIQVDGEELVGRSP-----GPNVAVNAKGSVIIG 228
DB 173 ARLVSEDPIDNGEWRVYALRGGRGSIQVDGEELVGRSP-----GPNVAVNAKGSVIIG 228

Db 990 AVLRSKAPIPLNVNVVTVNNGRKLKINKDELVSCEPSKSPKAPHTALNLKEAFYVG 1049
QY 229 GAPD---VATLTGGRFSSGIGTCVKNVLVLS 256
DB 1050 GAPDFNKPARAAG--IISGFTGAIGKLSLKS 1078

RESULT 6
AGRI_CHICK
ID AGRI_CHICK STANDARD; PRT; 1955 AA.
AC P31696;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Agtrin precursor.
GN AGRN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]_TaxID=9031;
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92232297; PubMed=1314620;
RA Ts'ao K.W.K., Ruegg M.A., Escher G., Kroeger S., McMahon U.J.;
RT "cDNA that encodes active agtrin.";
RL Neuron 8:677-689(1992).
RN [2]
RP ALTERNATIVE SPLICING.
RX MEDLINE=92232298; PubMed=1314621;
RA Ruegg M.A., Ts'ao K.W.K., Horton S.E., Kroeger S., Escher G.,
RA Gensch E.M., McMahon U.J.;
RT "The agtrin gene codes for a family of basal lamina proteins that
RT differ in function and distribution.";
RL Neuron 8:691-699(1992).
CC -|- FUNCTION: Component of the basal lamina that causes the
CC aggregation of acetylcholine receptors and acetylcholine-esterase
CC on the surface of muscle fibers of the neuromuscular junction.
CC -|- SUBUNIT: Binds to laminin.
CC -|- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
CC junction.
CC -|- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=3; Isoforms differ in
CC their acetylcholine receptor clustering activity;
CC Name=1;
CC IsoId=P31696-1; Sequence=Displayed;
CC Name=2; Synonyms=Agtrin-related protein 1;
CC IsoId=P31696-2; Sequence=VSP_001370;
CC Name=3; Synonyms=Agtrin-related protein 2;
CC IsoId=P31696-3; Sequence=VSP_001369, VSP_001370;
CC -|- PTM: Contains heparan sulfate chains as well as N-linked and O-
CC linked oligosaccharides (By similarity).
CC -|- SIMILARITY: Contains 9 Kazal-like domains.
CC -|- SIMILARITY: Contains 2 laminin EGF-like domains.
CC -|- SIMILARITY: Contains 4 EGF-like domains.
CC -|- SIMILARITY: Contains 1 SEA domain.
CC -|- SIMILARITY: Contains 3 laminin G-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M94271; AAA48585.1; --
CC EMBL; M97371; AAA48586.1; --
CC EMBL; M97372; --; NOT_ANNOTATED_CDS.
CC PIR; JH0591; AGCH.
CC HSP; P00740; LEDM.
```

DR InterPro; IPR004850; Agrin_NTA.
 DR InterPro; IPR00152; Asx_hydroxyl_S.
 DR InterPro; IPR008985; Cona_like_lec_gl.
 DR InterPro; IPR00742; EGF_2.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR003645; FOLN.
 DR InterPro; IPR002350; Kazal.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR00082; SEA_domain.
 DR InterPro; IPR008993; TIMP_like.
 DR Pfam; PF00008; EGF; 4.
 DR Pfam; PF00050; kazal; 9.
 DR Pfam; PF00053; laminin_EGF; 2.
 DR Pfam; PF00054; laminin_G; 3.
 DR Pfam; PF03146; NEA; 1.
 DR Pfam; PF01390; SEA; 1.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00180; EGF_Lam; 2.
 DR SMART; SM00274; FOLN; 8.
 DR SMART; SM00280; KAZAL; 9.
 DR SMART; SM00282; LamG; 3.
 DR SMART; SM00200; SEA; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00022; EGF_1; 6.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS00026; EGF_3; 4.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
 DR PROSITE; PS00024; SEA; 1.
 KW Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
 KW Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
 FT SIGNAL 1 38
 FT CHAIN 39 1955
 FT DOMAIN 54 126
 FT DOMAIN 130 201
 FT DOMAIN 202 273
 FT DOMAIN 276 344
 FT DOMAIN 350 418
 FT DOMAIN 419 483
 FT DOMAIN 484 548
 FT DOMAIN 551 633
 FT DOMAIN 675 728
 FT DOMAIN 729 775
 FT DOMAIN 781 851
 FT DOMAIN 1026 1148
 FT DOMAIN 1229 1265
 FT DOMAIN 1270 1445
 FT DOMAIN 1446 1483
 FT DOMAIN 1485 1522
 FT DOMAIN 1532 1713
 FT DOMAIN 1714 1752
 FT DOMAIN 1776 1952
 FT DOMAIN 856 995
 FT DOMAIN 1150 1219
 FT DOMAIN 86 105
 FT DOMAIN 94 126
 FT DOMAIN 160 180
 FT DOMAIN 169 201
 FT DOMAIN 233 252
 FT DOMAIN 241 273
 FT DOMAIN 304 323
 FT DOMAIN 312 344
 FT DOMAIN 378 397
 FT DOMAIN 386 418
 FT DOMAIN 443 462
 FT DOMAIN 451 483
 FT DOMAIN 507 527
 FT DOMAIN 516 548
 FT DOMAIN 592 612
 FT DOMAIN 601 633
 FT DOMAIN 675 687
 FT DISULFID 677 694

FT DISULFID 696 705 BY SIMILARITY.
 FT DISULFID 708 726 BY SIMILARITY.
 FT DISULFID 729 741 BY SIMILARITY.
 FT DISULFID 731 748 BY SIMILARITY.
 FT DISULFID 750 759 BY SIMILARITY.
 FT DISULFID 762 773 BY SIMILARITY.
 FT DISULFID 810 830 POTENTIAL.
 FT DISULFID 819 851 POTENTIAL.
 FT DISULFID 1233 1244 BY SIMILARITY.
 FT DISULFID 1238 1253 BY SIMILARITY.
 FT DISULFID 1255 1264 BY SIMILARITY.
 FT DISULFID 1450 1461 BY SIMILARITY.
 FT DISULFID 1455 1471 BY SIMILARITY.
 FT DISULFID 1473 1482 BY SIMILARITY.
 FT DISULFID 1489 1500 BY SIMILARITY.
 FT DISULFID 1494 1510 BY SIMILARITY.
 FT DISULFID 1512 1521 BY SIMILARITY.
 FT DISULFID 1718 1731 BY SIMILARITY.
 FT DISULFID 1725 1740 BY SIMILARITY.
 FT DISULFID 1742 1751 BY SIMILARITY.
 FT CARBOHYD 390 390 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 659 659 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 764 764 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 814 814 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPPLIC 1648 1651 Missing (in isoform 3).
 FT VARSPPLIC 1783 1793 Missing (in isoform 2 and isoform 3).
 FT CONFLICT 1129 1131 RTI > SIL (IN REF. 1; AAA48586).
 FT SEQUENCE 1955 AA; 211411 MW; B4DEB27C3422581 CRC64;
 Query Match 19.3%; Score 301.5; DB 1; Length 1955;
 Best Local Similarity 27.4%; Pred. No. 8.6e-14;
 Matches 87; Conservative 38; Mismatches 108; Indels 85; Gaps 13;
 QY 1 CERQPCQCATCPAGEVEFCLECRDGFGLCEHSENPCLRECLHGGTC-----OGT 55
 Db 1450 CHNPFCHGASCHVKAEMFHCCLHSYTGPTCADERNPCD-PTFCHISATCLVLPBGGA 1508
 QY 56 RCLCLPGFSGPRCQGGHGAESDWHLEGGGNDAPQGYGAYFHDGDLAPFGCHVFSRS 115
 Db 1509 MCACPMGREGFCER-----VTEQD-----HT 1530
 QY 116 LPEVPE-----TILEVETSTASGLLWQGVVEGAGGKDFI 153
 Db 1531 MPFLPEFNGFSYLELNGQLTLFLTCRQMSMEVFLAKSPGMIFYNGQKTD--GRG-DFV 1587
 QY 154 SLGLQGHVFRYQLGSGEARLVSDPDINDGEWHRVTLALREGRGSIQVDEELVSGRSP 213
 Db 1588 SLALHDGVLEYDLDLGAALVLRKEPVLNTWISVLLERSGRKGVMRINNGERVWGESP 1647
 QY 214 G-----PNVAVNAKGSVIYIGGAPDVATLT-GRGFSGGITGCVKNLVLHSAFGAPPPQPLD 268
 Db 1648 KSRKVPFAFLNLKEPPYVGGAPDFSKLAAAAIAISTSFYGAQVQIRISI-----KGVP---LLK 1700
 QY 269 LOH-RAQAGANT---RPC 282
 Db 1701 EQHRSVEISITFAHPC 1718

RESULT 7

FAT_DROME ID_FAT_DROME STANDARD; PRT; 5147 AA.
 AC P3350; Q9VQX5;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cadherin-related tumor suppressor precursor (Fat protein).
 GN FT OR CG3352.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92069752; PubMed=1959133;
 RA Mahoney P.A., Weber U., Onofrechuk P., Biessmann H., Bryant P.J.,
 RA Goodman C.S.;
 RT "The fat tumor suppressor gene in Drosophila encodes a novel member
 RT of the cadherin gene superfamily.";
 RL Cell 67:853-868(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Fendell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.-H.C., Blake R.G., Champe M., Pfeiffer B.D.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram P.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Borchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Flesher C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kallal B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirezka R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weisscock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -!- FUNCTION: Could function as a cell-adhesion protein. Acts as a
 CC tumor suppressor. Required for correct morphogenesis.
 CC -!- SIMILARITY: Contains 34 cadherin domains.
 CC -!- SIMILARITY: Contains 5 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 laminin G-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M80537; AAA28530.1; -;
 CC EMBL: AB003577; AAF51036.1; -;
 CC HSSP: P00740; LEDM.
 CC FlyBase: FBgn0001075; ft.
 CC GO: GO:0005887; C:intrigues to plasma membrane; NAS.
 CC GO: GO:0008014; F:calcium-dependent cell-cell adhesion; NAS.
 CC GO: GO:0016339; P:calcium-dependent cell-cell adhesion; NAS.
 CC GO: GO:0008283; P:cell proliferation; IMP.

DR GO: GO:000904; P:cellular morphogenesis during differentiation; IMP.
 DR GO: GO:0045317; P:equator specification; IMP.
 DR GO: GO:0045198; P:establishment of epithelial cell polarity; IMP.
 DR GO: GO:0007446; P:imaginal disc growth; IMP.
 DR GO: GO:0018149; P:protein-protein cross-linking; IPI.
 DR InterPro: IPR002126; Cadherin.
 DR InterPro: IPR008985; ConA_like_lect_gl.
 DR InterPro: IPR000742; EGF_2.
 DR InterPro: IPR006209; EGF-like.
 DR InterPro: IPR006210; EGF.
 DR InterPro: IPR001791; Laminin G.
 DR Pfam: PF00028; cadherin; 34.
 DR Pfam: PF00008; EGF; 4.
 DR Pfam: PF00054; laminin G; 2.
 DR PRINTS: PR00205; CADHERIN.
 DR SMART: SM00112; CA; 34.
 DR SMART: SM00181; EGF; 4.
 DR SMART: SM00282; LamG; 2.
 DR PROSITE: PS00232; CADHERIN_1; 22.
 DR PROSITE: PS00268; CADHERIN_2; 34.
 DR PROSITE: PS00022; EGF_1; 4.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS00036; EGF_3; 4.
 DR PROSITE: PS00035; LAM_G_DOMAIN; 2.
 DR Cell adhesion; Signal; Transmembrane; Glycoprotein; Calcium-binding;
 KW Repeat; EGF-like domain.
 FT SIGNAL 1 35
 FT CHAIN 36 5147
 FT DOMAIN 36 5147
 FT TRANSMEM 4584 4609
 FT DOMAIN 4610 5147
 FT DOMAIN 36 156
 FT DOMAIN 157 270
 FT DOMAIN 271 382
 FT DOMAIN 383 494
 FT DOMAIN 495 599
 FT DOMAIN 600 708
 FT DOMAIN 709 820
 FT DOMAIN 821 942
 FT DOMAIN 943 1049
 FT DOMAIN 1050 1153
 FT DOMAIN 1154 1278
 FT DOMAIN 1279 1384
 FT DOMAIN 1385 1489
 FT DOMAIN 1490 1601
 FT DOMAIN 1602 1713
 FT DOMAIN 1714 1823
 FT DOMAIN 1824 1922
 FT DOMAIN 1923 2027
 FT DOMAIN 2028 2167
 FT DOMAIN 2168 2278
 FT DOMAIN 2279 2385
 FT DOMAIN 2386 2491
 FT DOMAIN 2492 2596
 FT DOMAIN 2597 2703
 FT DOMAIN 2704 2810
 FT DOMAIN 2811 2913
 FT DOMAIN 2914 3013
 FT DOMAIN 3014 3124
 FT DOMAIN 3125 3229
 FT DOMAIN 3230 3334
 FT DOMAIN 3335 3439
 FT DOMAIN 3440 3545
 FT DOMAIN 3546 3651
 FT DOMAIN 3652 3756
 FT DOMAIN 3950 4011
 FT DOMAIN 4013 4049
 FT DOMAIN 4052 4090
 FT DOMAIN 4092 4128
 FT DOMAIN 4129 4320
 FT DOMAIN 4321 4362
 FT DOMAIN 4402 4569
 FT DISULFID 3954 3966
 POTENTIAL.
 CADHERIN-RELATED TUMOR SUPPRESSOR.
 EXTRACELLULAR (POTENTIAL).
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).
 CADHERIN 1.
 CADHERIN 2.
 CADHERIN 3.
 CADHERIN 4.
 CADHERIN 5.
 CADHERIN 6.
 CADHERIN 7.
 CADHERIN 8.
 CADHERIN 9.
 CADHERIN 10.
 CADHERIN 11.
 CADHERIN 12.
 CADHERIN 13.
 CADHERIN 14.
 CADHERIN 15.
 CADHERIN 16.
 CADHERIN 17.
 CADHERIN 18.
 CADHERIN 19.
 CADHERIN 20.
 CADHERIN 21.
 CADHERIN 22.
 CADHERIN 23.
 CADHERIN 24.
 CADHERIN 25.
 CADHERIN 26.
 CADHERIN 27.
 CADHERIN 28.
 CADHERIN 29.
 CADHERIN 30.
 CADHERIN 31.
 CADHERIN 32.
 CADHERIN 33.
 CADHERIN 34.
 EGF-LIKE 1.
 EGF-LIKE 2.
 EGF-LIKE 3.
 EGF-LIKE 4.
 LAMININ G-LIKE 1.
 EGF-LIKE 5.
 LAMININ G-LIKE 2.
 BY SIMILARITY.

```
FT DISULFID 3960 3999 BY SIMILARITY.
FT DISULFID 4001 4010 BY SIMILARITY.
FT DISULFID 4017 4028 BY SIMILARITY.
FT DISULFID 4039 4047 BY SIMILARITY.
FT DISULFID 4056 4067 BY SIMILARITY.
FT DISULFID 4061 4078 BY SIMILARITY.
FT DISULFID 4080 4089 BY SIMILARITY.
FT DISULFID 4096 4107 BY SIMILARITY.
FT DISULFID 4101 4116 BY SIMILARITY.
FT DISULFID 4118 4127 BY SIMILARITY.
FT DISULFID 4325 4341 BY SIMILARITY.
FT DISULFID 4334 4350 BY SIMILARITY.
FT DISULFID 4352 4361 BY SIMILARITY.
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 461 461 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 605 605 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 1155 1155 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 1367 1367 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 1458 1458 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 1751 1751 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 1831 1831 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 1880 1880 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 2080 2080 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 2171 2171 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 2247 2247 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 2290 2290 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 2437 2437 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 2581 2581 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 2799 2799 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 2920 2920 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 2946 2946 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 2967 2967 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 3167 3167 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 3303 3303 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 3386 3386 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 3389 3389 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)
FT CARBOHYD 3525 3525 N-LINKED (GLCNAC. . .) (POTENTIAL. . .)

Query Match 17.2%; Score 269; DB 1; Length 5147;
Best Local Similarity 29.7%; Pred. No. 5.1e-11;
Matches 94; Conservative 39; Mismatches 102; Indels 82; Gaps 19;

Qy 1 CERQPCOHGATCM--PAGEVEFQCLCRDGFKGDLCEHENPCQLRFPCLHGTC-----QG 54
Db 4056 CYSKPCRNCGSCORSPGSSYF-CLCRPGRNQCSVSQSCR-PNPLHGLGCVSLKPG 4113
Qy 55 TRCLCLFPGSPRCCQSGHGHIAESDWHLRGSGNDAPGOYGAYFHDGFLAPPGHVFSR 114
Db 4114 YKCNCTEGRYGRHCERFS-----YQ--FQPLSVMTFP----- 4143
Qy 115 SLPEVPTIELEVRTSTAGLLIWO-GVEVGEAGQKDFISLGLQCHLVFRYQLGSGEA 173
Db 4144 ALDVTNTDISIVFATTPNLSLLLYNGM---QSGGRSDFLAIELVHGRAYF-----SSGGA 4196
Qy 174 R-----LVSEDPINDGSEHWRTALREGRGSIQV-----DGE---ELVGRSP-----G 214
Db 4197 RTALSTVIAGNLADGGWHKVTATNGRVNLSVAKCAUSDVCTCLPDSSCYADEVG 4256
Qy 215 PNVAVN-AKGSVIYG---APDVATLTGGRFFSGITGCVKNVLVHSGARCPAPPQPLDL- 269
Db 4257 PVGTLNFKOPLMIGLSSADPILRPGQVHSDDLVGC-----LHSHVHGG---RALNLS 4308
Qy 270 ----QHEAQAQANTRPC 282
Db 4309 LPLOQKILAGCNKQAC 4325
```

```
RESULT 8
SLT2 HUMAN
ID SLT2 HUMAN STANDARD; PRT; 1529 AA.
AC O94813; O95710; Q9YSQ7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Slit homolog 2 protein precursor (h-Slit-2).
GN SLT2 OR SLIT2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
RP TISSUE=Fetal lung;
RX MEDLINE=99033071; PubMed=9813312;
RA Itoh A., Miyabayashi T., Ohno M., Sakano S.;
RT "Cloning and expressions of three mammalian homologues of Drosophila
RT slit suggest possible roles for slit in the formation and maintenance
RT of the nervous system";
RL Brain Res. Mol. Brain Res. 62:175-186(1998).
RN [2]
SEQUENCE FROM N.A. (ISOFORM 3), AND TISSUE SPECIFICITY.
RP TISSUE=Fetal brain, and Fetal kidney;
RX MEDLINE=99279238; PubMed=10349621;
RA Holmes G.P., Negus K., Burridge L., Raman S., Algar E., Yamada T.,
RA Little M.H.;
RT "Distinct but overlapping expression patterns of two vertebrate slit
RT homologs implies functional roles in CNS development and
RT organogenesis.";
RL Mech. Dev. 79:57-72(1998).
RN [3]
SEQUENCE FROM N.A. (ISOFORM 2), SEQUENCE OF 1122-1129, FUNCTION,
RP SUBUNIT AND SUBCELLULAR LOCATION.
RX TISSUE=Fetal brain;
RX MEDLINE=99200391; PubMed=10102268;
RA Brose K., Bland K.S., Wang K.H., Arnett D., Henzel W., Goodman C.S.,
RA Tessier-Lavigne M., Kidd T.;
RT "Slit proteins bind Robo receptors and have an evolutionarily
RT conserved role in repulsive axon guidance.";
RL Cell 96:795-806(1999).
CC -!- FUNCTION: Plays a role in axon guidance as a repulsive ligand for
CC Robo receptors preventing inappropriate midline crossing.
CC -!- SUBUNIT: Binds robo proteins with high affinity.
CC -!- SUBCELLULAR LOCATION: Secreted. The C-terminal cleavage protein is
CC more diffusible than the larger N-terminal protein that is more
CC tightly cell associated.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC Name=2;
CC Name=3;
CC IsoId=O94813-1; Sequence=VSP_050035, VSP_050036;
CC IsoId=O94813-2; Sequence=VSP_050035, VSP_050036;
CC IsoId=O94813-3; Sequence=VSP_050036;
CC -!- TISSUE SPECIFICITY: Fetal lung and kidney, and adult spinal cord.
CC Weak expression in adult adrenal gland, thyroid, trachea and other
CC tissues examined.
CC -!- SIMILARITY: Contains 7 EGF-like domains.
CC -!- SIMILARITY: Contains 23 leucine-rich (LRR) repeats.
CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
CC -!- SIMILARITY: Contains 1 laminin G-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB017168; BAA35185.1; -
CC DR
```


QY 92 -----PGQYGA-----YFHDGFLAFPGHVSRLPEVPT-IELEVRISTAS 133
 Db 1141 NEPICQCLFGYQSGKCEKLVSNFINKESYLQIP-----SAKVRPQNTITQIATDEDS 1194
 QY 134 GLLLMQGVGEAGQGGDFISLQDQHLVFRYQLGSGEARLV-SEDPINDGWEHRTAL 192
 Db 1195 GILLYKG-----DKDHIAVELYGRVRSYDTGSHPASAIYSVETINDGNPHIVELL 1246
 QY 193 REGRRGSTQVQG--BELVSGRSPGNVAVNAKGVYIGGAP---DVATL---TGRFSSGI 245
 Db 1247 ALDQSLSLVDGNGPKIITNLS--KQSTLNPDSPLYVCGMPGKSNVSLRQAPGQNGTSF 1304
 QY 246 TGCYKVLNLS 256
 Db 1305 HGCIRNLVINS 1315

RESULT 9
 FAT2 DROME
 ID FAT2 DROME STANDARD; PRT; 4705 AA.
 AC Q9VW71, Q9VW51,
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Putative fat-like cadherin-related tumor suppressor homolog precursor.
 DE
 GN FAT2 OR CG7749.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=2019606; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.M., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."

Science 287:2185-2195(2000).
 [2]
 RN REVISIONS.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.K., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review."
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RN [3]
 SEQUENCE OF 3837-4705 FROM N.A.
 RC STRAIN=Berkley; TISSUE=Ovary;
 RX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celniker S.E.,
 RT "A Drosophila full-length cDNA resource."
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -!- SIMILARITY: Contains 34 cadherin domains.
 CC -!- SIMILARITY: Contains 5 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 laminin G-like domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AE003515; AAF49078.2; -
 DR EMBL; AY060955; AAL28503.1; ALT_INIT.
 DR EMBL; AY118666; AAMS0035.1; ALT_INIT.
 DR HSSP; P15116; INCI.
 DR FlyBase; FBgn0036930; fat2.
 DR GO; GO:0005887; C:integral to plasma membrane; ISS.
 DR GO; GO:0008004; F:calcium-dependent cell adhesion molecule ac. .; ISS.
 DR GO; GO:0016339; F:calcium-dependent cell-cell adhesion; ISS.
 DR InterPro; IPR000152; Asx hydroxyl_S.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR008985; ConA like_lec_gi.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00028; cadherin; 31.
 DR Pfam; PF00008; EGF; 5.
 DR Pfam; PF00054; laminin_G; 1.
 DR PRINTS; PR00205; CADHERIN.
 DR SMART; SM00112; CA; 34.
 DR SMART; SM00181; EGF; 6.
 DR SMART; SM00282; LamG; 1.
 DR PROSITE; PS00010; ASX HYDROXYL; 1.
 DR PROSITE; PS00232; CADHERIN_1; 18.
 DR PROSITE; PS02268; CADHERIN_2; 34.
 DR PROSITE; PS00022; EGF_1; 5.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS50026; EGF_3; 5.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS50025; LAM_G DOMAIN; 1.
 KW Hypothetical protein; Cell adhesion; Signal; Glycoprotein;
 KW Transmembrane; Calcium; Calcium-binding; Repeat; EGF-like domain.
 FT SIGNAL 1 35
 FT CHAIN 36 4705 PUTATIVE FAT-LIKE CADHERIN-RELATED TUMOR
 FT DOMAIN 36 1647 SUPPRESSOR HOMOLOG.
 FT TRANSMEM 1648 1668 EXTRACELLULAR (POTENTIAL).
 FT

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Glaesner W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaesner K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laško P., Lei Y., Levitsky A.C., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobaraj C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.M., Nelson K.A., Nixon K.A., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zavari J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*,"
RL Science 287:2185-2195(2000).
RN [4]
RP REVISIONS.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradscky P., Huang Y., Kaminler J.S., Millburn G.H., Prochuk S.B.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Battencourt B.E., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.,
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RT systematic review,"
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
CC -1- FUNCTION: Involved in the fz signaling pathway that controls wing
CC tissue polarity. Also mediates homophilic cell adhesion. May play
CC a role in initiating prehair morphogenesis. May play a critical
CC role in tissue polarity and in formation of normal dendrite
CC fields.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: In the pupal wing, expressed at relatively
CC even levels in all regions. Abundant in 6-9 hour embryos.
CC Expressed at higher levels in pupae than larvae.
CC -1- DEVELOPMENTAL STAGE: At 12 hours after puparium formation (apf),
CC is concentrated at proximal and distal cell boundaries with little
CC or no expression at anterior and posterior boundaries. When
CC prehairsts emerge at 30-36 hours apf, expression becomes evenly
CC distributed again along the whole cell boundary.
CC -1- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
CC -1- SIMILARITY: Contains 8 cadherin domains.
CC -1- SIMILARITY: Contains 4 EGF-like domains.
CC -1- SIMILARITY: Contains 2 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 GPS domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF172329; AAF02618.1; -
CC EMBL; AB028498; BAA84069.1; -
CC EMBL; AE003828; AAF58763.3; -
CC HSSP; P08709; IBF9
CC FlyBase; FBgn0024836; stan.
CC GO; GO:0016021; C:integral to membrane; NAS.

DR GO; GO:0008014; P:calcium-dependent cell adhesion molecule ac. .; IMP.
DR GO; GO:0005057; F:receptor signaling protein activity; IMP.
DR GO; GO:0016358; P:dendrite morphogenesis; IMP.
DR GO; GO:0007222; P:fzrizzled signaling pathway; IMP.
DR GO; GO:0007367; P:segment polarity determination; IMP.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR008985; ConA-like_lect_gli.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Like.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; hormn_receptor.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000203; PKD_cys-rich.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF00028; cadherin; 8.
DR Pfam; PF00008; EGF; 3.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF00053; laminin_EGF; 1.
DR Pfam; PF00054; laminin_G; 1.
DR PRINTS; PR00205; CADHERIN.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR00249; GPCRSECRETIN.
DR SMART; SM00112; CA; 8.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00303; GPS_1.
DR SMART; SM00282; LamG; 2.
DR PROSITE; PS00232; CADHERIN_1; 6.
DR PROSITE; PS0268; CADHERIN_2; 8.
DR PROSITE; PS00022; EGF_1; 4.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS00026; EGF_3; 3.
DR PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
DR PROSITE; PS0227; G_PROTEIN_RECEP_F2_3; 1.
DR PROSITE; PS0261; G_PROTEIN_RECEP_F2_4; 1.
DR PROSITE; PS0221; GPS; 1.
DR PROSITE; PS00025; LAM_G_DOMAIN; 2.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW Cell adhesion: Developmental protein; G-protein coupled receptor;
KW Calcium-binding; Repeat; Signal; Transmembrane; EGF-like domain;
KW Laminin EGF-like domain; Glycoprotein.
FT SIGNAL 1 29
FT CHAIN 30 3579
FT PROTOCADHERIN-LIKE WING POLARITY PROTEIN
FT STAN.
FT DOMAIN 30 2816
FT TRANSMEM 2817 2837
FT DOMAIN 2838 2845
FT TRANSMEM 2846 2866
FT DOMAIN 2867 2883
FT TRANSMEM 2884 2904
FT DOMAIN 2905 2919
FT TRANSMEM 2920 2940
FT DOMAIN 2941 2959
FT TRANSMEM 2960 2980
FT DOMAIN 2981 3000
FT TRANSMEM 3001 3021
FT DOMAIN 3022 3031
FT TRANSMEM 3032 3052
FT DOMAIN 3053 3079
FT TRANSMEM 3080 3097
FT DOMAIN 3098 3113
FT TRANSMEM 3114 3123
FT DOMAIN 3124 3158
FT TRANSMEM 3159 3175
FT DOMAIN 3176 3192

```
FT DOMAIN 1796 1963 LAMININ G-LIKE 2.
FT DOMAIN 1965 2000 EGF-LIKE 3 CALCIUM-BINDING.
FT DOMAIN 2091 2126 EGF-LIKE 4 CALCIUM-BINDING.
FT DOMAIN 2744 2802 GPS.
FT DOMAIN 140 143 POLY-HIS.
FT DOMAIN 155 159 POLY-ARG.

Query Match 14.4%; Score 225; DB 1; Length 3579;
Best Local Similarity 22.4%; Pred. No. 4.7e-08;
Matches 83; Conservative 52; Mismatches 120; Indels 116; Gaps 17;

QY 1 CERQCOHGATCMP-----AGEY-----EFQCLCRDGFKGD-----LCEH 35
DB 1422 CVREPLNFEELTVLKFNASEFTSDTLVPRTPYVNTFACSGPEGFTGSKHYLCDT 1481
QY 36 ENPCLARPCPLHGTC-----QFTCLCLPFGSPRCQOQSGH-----GTAESDWHLEG 85
DB 1482 EVDLC-YSDPCQNGTCVRRREGYTCVCPSTHTGNCETGVGHLPDPCPSETCEGSLCLS 1540
QY 86 SGNDAPGOYGA-----YFDDGFLAPPG-----HVFSRSLPEVPETIELEVRTSTASG 134
DB 1541 NYFSSQPPPTATCELARAFGRNSFLTFSFKQRHFN-----LKLRFATVQENG 1591
QY 135 LLLWQVEVEGAGQKDPISLQDGHVLPYQLSGEARL--VSEDPINDGEWHRYTAL 192
DB 1592 LLLYN-----GRYNELHDFIALEIHEGHVSFSLGDHSERISVIOEAKVSDGKHQVVEW 1647
QY 193 REGRGSIQVDCGE-----LVSGR-----SPGNVAVNAKGSVYI 227
DB 1648 YLNRSVTLVLDNCDTAIALSGQLGRWSCANRTTLKDKRCSLLTETCHRELDLTGPLOV 1707
QY 228 GGAPDVAT---ITGGRFSSGITGCVKNLVL-----HSARGAPPPDPLDLOH 271
DB 1708 GGLPRIPAFHPTNRDF-----VGCISDLRIDDFVLDLSYVADNGTLGACQPKAPL----- 1759
QY 272 RAQAGANTRPC 282
DB 1760 -----CQSEPC 1765

RESULT 11
LMA2 HUMAN
ID LMA2 HUMAN STANDARD; PRT; 3110 AA.
AC P24043; Q14736; Q93022;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Laminin alpha-2 chain precursor (Laminin M chain) (Merosin heavy
DE chain).
GN LMA2 OR LAMM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=94124633; PubMed=8294519;
RA Vuolteenaho R., Nissinen M., Sainio K., Byers M., Eddy R.,
RA Hirvonen H., Shows T.B., Sariola H., Engvall E., Tryggvason K.;
RT "Human laminin M chain (merosin): complete primary structure,
RT chromosomal assignment, and expression of the M and A chain in human
RT fetal tissues."
RL J. Cell Biol. 124:381-394(1994).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=97066955; PubMed=8910357;
RA Zhang X., Vuolteenaho R., Tryggvason K.;
RT "Structure of the human laminin alpha2-chain gene (LAMA2), which is
RT affected in congenital muscular dystrophy."
RL J. Biol. Chem. 271:27664-27669(1996).
RN [3]
RN SEQUENCE OF 1981-3110 FROM N.A., AND PARTIAL SEQUENCE.
```

```
RC TISSUE=Placenta;
RX MEDLINE=90238994; PubMed=2185464;
RA Ehrig K., Leivo I., Argraves W.S., Ruoslahti E., Engvall E.;
RT "Merosin, a tissue-specific basement membrane protein, is a
RT laminin-like protein."
RL Proc. Natl. Acad. Sci. U.S.A. 87:3264-3268(1990).
RN [4]
RN VARIANTS GLN-545; HIS-619; LEU-919; HIS-2586 AND LYS-2614.
RA Panicker S.G., Mendell J.T., Chen L., Feng B., Sahenk Z.,
RA Marzluf G.A., Amato A.A., Mendell J.R.;
RA Marzluf G.A., Amato A.A., Mendell J.R.;
RT "Novel single base polymorphisms and rare sequence variants in
RT the laminin 2-chain coding region detected by RNA/SSCP analysis."
RL Hum. Mutat. 13:174-174(1999).
RN [5]
RN ERRATUM.
RA Panicker S.G., Mendell J.T., Chen L., Feng B., Sahenk Z.,
RA Marzluf G.A., Amato A.A., Mendell J.R.;
RL Hum. Mutat. 13:340-340(1999).
RN [6]
RN VARIANT MDC1A PRO-2564.
RX MEDLINE=21476011; PubMed=11591858;
RA He Y., Jones K.J., Vignier N., Morgan G., Chevallay M., Barois A.,
RA Estournet-Mathiaud B., Hori H., Mizuta T., Tome F.M.S., North K.N.,
RA Guicheney P.;
RT "Congenital muscular dystrophy with primary partial laminin alpha-2
RT chain deficiency: molecular study."
RL Hum. Mutat. 21:103-111(2003).
RN [7]
RN VARIANTS MDC1A TYR-527 AND ARG-862.
RX MEDLINE=22439669; PubMed=12552556;
RA Tezak Z., Prandini P., Boccardo M., Marin A., Devaney J., Marino M.,
RA Fanin M., Trevisan C.P., Park J., Tyson W., Finkel R., Garcia C.,
RA Angelini C., Hoffman E.P., Pegoraro E.;
RT "Clinical and molecular study in congenital muscular dystrophy with
RT partial laminin alpha-2 (LAMA2) deficiency."
RL Hum. Mutat. 21:103-111(2003).
CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC is thought to mediate the attachment, migration and organization
CC of cells into tissues during embryonic development by interacting
CC with other extracellular matrix components.
CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC different polypeptide chains (alpha, beta, gamma), which are bound
CC to each other by disulfide bonds into a cross-shaped molecule
CC comprising one long and three short arms with globules at each
CC end. The alpha-2 chain is a subunit of laminin-2 (merosin) and
CC laminin-4 (S-merosin). Extracellular; found in the basement
CC membranes (major component).
CC -1- TISSUE SPECIFICITY: Placenta, striated muscle, peripheral nerve,
CC cardiac muscle, pancreas, lung, spleen, kidney, adrenal gland,
CC skin, testis, meninges, choroid plexus, and some other regions of
CC the brain; not in liver, thymus and bone.
CC -1- DOMAIN: The alpha-helical domains I and II are thought to interact
CC with other laminin chains to form a coiled coil structure.
CC -1- DOMAIN: Domains VI, IV and G are globular.
CC -1- DISEASE: Defects in LAMA2 are the cause of merosin-deficient
CC congenital muscular dystrophy type 1A (MDC1A) [MIM:607855]. MDC1A
CC is characterized by difficulty walking, hypotonia, proximal
CC weakness, hyporeflexia, and white matter hypodensity on MRI.
CC -1- SIMILARITY: Contains 1 laminin N-terminal domain.
CC -1- SIMILARITY: Contains 17 laminin EGF-like domains.
CC -1- SIMILARITY: Contains 2 laminin IV domains.
CC -1- SIMILARITY: Contains 5 laminin G-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z26653; CAA81394.1; -.
```

```
DR EMBL; U66796; AAB18388.1; JOINED.
DR EMBL; U66733; AAB18388.1; JOINED.
DR EMBL; U66734; AAB18388.1; JOINED.
DR EMBL; U66735; AAB18388.1; JOINED.
DR EMBL; U66736; AAB18388.1; JOINED.
DR EMBL; U66737; AAB18388.1; JOINED.
DR EMBL; U66738; AAB18388.1; JOINED.
DR EMBL; U66739; AAB18388.1; JOINED.
DR EMBL; U66740; AAB18388.1; JOINED.
DR EMBL; U66741; AAB18388.1; JOINED.
DR EMBL; U66742; AAB18388.1; JOINED.
DR EMBL; U66743; AAB18388.1; JOINED.
DR EMBL; U66744; AAB18388.1; JOINED.
DR EMBL; U66745; AAB18388.1; JOINED.
DR EMBL; U66746; AAB18388.1; JOINED.
DR EMBL; U66747; AAB18388.1; JOINED.
DR EMBL; U66748; AAB18388.1; JOINED.
DR EMBL; U66749; AAB18388.1; JOINED.
DR EMBL; U66750; AAB18388.1; JOINED.
DR EMBL; U66751; AAB18388.1; JOINED.
DR EMBL; U66752; AAB18388.1; JOINED.
DR EMBL; U66753; AAB18388.1; JOINED.
DR EMBL; U66754; AAB18388.1; JOINED.
DR EMBL; U66755; AAB18388.1; JOINED.
DR EMBL; U66756; AAB18388.1; JOINED.
DR EMBL; U66757; AAB18388.1; JOINED.
DR EMBL; U66758; AAB18388.1; JOINED.
DR EMBL; U66759; AAB18388.1; JOINED.
DR EMBL; U66760; AAB18388.1; JOINED.
DR EMBL; U66761; AAB18388.1; JOINED.
DR EMBL; U66762; AAB18388.1; JOINED.
DR EMBL; U66763; AAB18388.1; JOINED.
DR EMBL; U66764; AAB18388.1; JOINED.
DR EMBL; U66765; AAB18388.1; JOINED.
DR EMBL; U66766; AAB18388.1; JOINED.
DR EMBL; U66768; AAB18388.1; JOINED.
DR EMBL; U66769; AAB18388.1; JOINED.
DR EMBL; U66770; AAB18388.1; JOINED.
DR EMBL; U66771; AAB18388.1; JOINED.
DR EMBL; U66772; AAB18388.1; JOINED.
DR EMBL; U66773; AAB18388.1; JOINED.
DR EMBL; U66774; AAB18388.1; JOINED.
DR EMBL; U66775; AAB18388.1; JOINED.
DR EMBL; U66776; AAB18388.1; JOINED.
DR EMBL; U66777; AAB18388.1; JOINED.
DR EMBL; U66778; AAB18388.1; JOINED.
DR EMBL; U66779; AAB18388.1; JOINED.
DR EMBL; U66780; AAB18388.1; JOINED.
DR EMBL; U66781; AAB18388.1; JOINED.
DR EMBL; U66782; AAB18388.1; JOINED.
DR EMBL; U66783; AAB18388.1; JOINED.
DR EMBL; U66784; AAB18388.1; JOINED.
DR EMBL; U66785; AAB18388.1; JOINED.
DR EMBL; U66786; AAB18388.1; JOINED.
DR EMBL; U66787; AAB18388.1; JOINED.
DR EMBL; U66788; AAB18388.1; JOINED.
DR EMBL; U66789; AAB18388.1; JOINED.
DR EMBL; U66790; AAB18388.1; JOINED.
DR EMBL; U66791; AAB18388.1; JOINED.
DR EMBL; U66792; AAB18388.1; JOINED.
DR EMBL; U66793; AAB18388.1; JOINED.
DR EMBL; U66794; AAB18388.1; JOINED.
DR EMBL; U66795; AAB18388.1; JOINED.
DR EMBL; M59832; AAB18388.1; JOINED.
DR PIR; FX0082; M59832.
DR HSSP; Q60675; 1Q00.
DR Genew; HGNC:6482; LAMA2.
DR MIM; 156225; -.
DR MIM; 607855; -.
DR GO; GO:0005604; C:basement membrane; TAS.
DR GO; GO:0005198; F:structural molecule activity; TAS.
DR GO; GO:0007517; P:muscle development; TAS.
DR InterPro; IPR00985; CoA_like_lec_gl.
DR InterPro; IPR006209; EGF_like.

DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR008211; LamNT.
DR Pfam; PF00052; laminin_B_2.
DR Pfam; PF00053; laminin_EGF_14.
DR Pfam; PF00054; laminin_G_5.
DR Pfam; PF00055; laminin_Nterm_1.
DR PRINTS; PR00011; EGF_LAMININ.
DR ProDom; PD003031; Laminin_B_1.
DR SMART; SM00180; EGF_Lam_15.
DR SMART; SM00281; LamB_2.
DR SMART; SM00282; LamG_5.
DR SMART; SM00136; LamNT_1.
DR PROSITE; PS00022; EGF_1_11.
DR PROSITE; PS01186; EGF_2_3.
DR PROSITE; PS01248; LAMININ_TYPE_EGF_14.
DR PROSITE; PS00025; LAM_G_DOMAIN_5.
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal; Polymorphism;
KW Disease mutation.
FT SIGNAL 1 22 POTENTIAL.

Query Match 14.3%; Score 223.5; DB 1; Length 3110;
Best Local Similarity 36.4%; Pred. No. 5.1e-08;
Matches 59; Conservative 22; Mismatches 68; Indels 13; Gaps 5;

QY 122 TIEVTRTSTAGLLMVGVEVGEAGQKDFISLQDCHLVFRYQLGSGEARLVSEDP 181
DB 2788 TIEVTRTEAESGLLFYNA-----AINHADPATVQLRNLGPFYSYDLGSGDTHMPTKI 2842
QY 182 NDGEWHRTALREGRGSIQVDGELVSGRSPGNVA--VNAKGSVIYGGAP-DVATLTG 238
DB 2843 NDGQWHEKIKMSKQEGILYVDG---ASNRTISPKADILDVWGLYVGLPINTYTRRI 2899
QY 239 GRFSSGITGVNVLVHSARGAPPQPLDLQHRQAQANTR 280
DB 2900 GPVTYSIDGCVRN--LHMAEAPADLEQPTSFHVGTCEPANAQ 2939

RESULT 12
NX3A_RAT
ID NX3A_RAT STANDARD; PRT; 1578 AA.
AC Q07310; Q07280; Q07311; Q07312; Q07313; Q07314;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neurexin 3-alpha precursor (Neurexin III-alpha).
GN NRXN3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]_TaxID=10116;
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=93342001; PubMed=8341647;
RA Ushkaryov Y.A., Suedhof T.C.;
RT "Neurexin III alpha: extensive alternative splicing generates
RT membrane-bound and soluble forms.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:6410-6414 (1993).
RN [2]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain;
RX MEDLINE=95209856; PubMed=7695896;
RA Ullrich B., Ushkaryov Y.A., Suedhof T.C.;
RT "Cartography of neurexins: more than 1000 isoforms generated by
RT alternative splicing and expressed in distinct subsets of neurons.";
RL Neuron 14:497-507 (1995).
RN [3]
RP SEQUENCE OF N-TERMINUS, AND INTERACTION WITH NEUREXOPHILIN 1.
RX MEDLINE=99074239; PubMed=9856994;
```


Alpha 5H)
/FTId=VSP_003532
SEQUENCE 1578 AA; 173993 MW; 936CF8529143D0C7 CRC64;
Query Match 13.7%; Score 214; DB 1; Length 1578;
Best Local Similarity 26.1%; Pred. No. 1.2e-07;
Matches 73; Conservative 38; Mismatches 121; Indels 48; Gaps 10;
QY 27 GPKGDLCE-----HEENPCQLREPLHGVC---QGTRCLC-LPGF 63
DB 171 GPKGLMLDKVGNSEPRLLSQSVLAEGPCGER-PCENGGICFLLDGHTCDCTTGY 229
QY 64 SGPRCQSGHGIAESDWHLEGGGNDAPQGYGAYFHDDGLAPPGHVFSRSLPEVPETI 123
DB 230 GGTLCSEDSVQCPGLSHLMSEQGRKXARENVATFRGSYLSV--DLQNPICSSSEI 287
QY 124 ELEVRTSTASGLLWQGVGEACQKDFISLGLDGLHVFYQLSGEARELVSEDP--181
DB 288 TLSFKTQWRGLL-----HTKSADYVNLKOGAVSLVNLGSGAFEAIVE-FVNG 339
QY 182 --NDGEWHRYTALREGRRGSIQVDEGLVSGRPGPNVAVNAKGVVYIGAPDVATLTGG 239
DB 340 KENDNAWHDVKNVNTLAQVIVSDGILITTYQEDYTMLGSDSSVYGFSPSTADLPSS 399
QY 240 RFSSGITGCVNLV--HSARPGAPPPQPLDLOHRAQGA 277
DB 400 PVSNFMGCLKEVYKNDIR-----LESLRLARIGA 431
RESULT 13
CLR2_HUMAN STANDARD; PRT; 2923 AA.
AC QHRCU4; Q92566;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cadherin EGF LAG seven-pass G-type receptor 2 precursor (Epidermal
DE growth factor-like 2) (Multiple epidermal growth factor-like domains
DE 3) (Flamingo 1).
GN CELSR2 OR CDHF10 OR EGFL2 OR MEGF3 OR KIAA0279.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=20363102; PubMed=10907856;
RA Vincent J.B., Skaug J., Scherer S.W.;
RT "The human homologue of flamingo, EGFL2, encodes a brain-expressed
RT large cadherin-like protein with epidermal growth factor-like domains,
RT and maps to chromosome 1p13.3-p21.1";
RL DNA Res. 7:233-235(2000).
[2]
SEQUENCE FROM N.A.
RA Suwa M., Sato T., Okouchi I., Arita M., Futami K., Matsumoto S.,
RA Tetsumi S., Aburatani H., Asai K., Akiyama Y.;
RT "Genome-wide discovery and analysis of human seven transmembrane helix
RT receptor genes";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE OF 516-2923 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayashi Y.,
RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain";
RL DNA Res. 3:321-329(1996).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Highest expression in brain and testis.

CC -1- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
CC -1- SIMILARITY: Contains 9 cadherin domains.
CC -1- SIMILARITY: Contains 8 EGF-like domains.
CC -1- SIMILARITY: Contains 2 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 GPS domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; AF234887; AAC00080.1; --
CC EMBL; AB065955; BAC06168.1; --
CC EMBL; D87469; BAA13407.1; --
CC HSSP; P15116; INCU.
CC Genew; HGNC:3231; CELSR2.
CC MIM; 604265; --
CC GO; GO:0016021; C:integral to membrane; NAS.
CC GO; GO:0004930; F:G-protein coupled receptor activity; NAS.
CC GO; GO:0007186; F:G-protein coupled receptor protein signalin. .; NAS.
CC InterPro; IPR00152; Asx_hydroxyl_S.
CC InterPro; IPR002126; Cadherin.
CC InterPro; IPR000742; EGF_2.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR006209; EGF-like.
CC InterPro; IPR000832; GPCR_secretin.
CC InterPro; IPR001879; hormn_receptor.
CC InterPro; IPR002049; Laminin_EGF.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR000203; PKD_cys_rich.
CC Pfam; PF00002; 7tm_2; 1.
CC Pfam; PF00028; cadherin; 8.
CC Pfam; PF00008; EGF; 5.
CC Pfam; PF01825; GPS; 1.
CC Pfam; PF02793; HRM; 1.
CC Pfam; PF00053; laminin_EGF; 1.
CC Pfam; PF00054; laminin_G; 1.
CC PRINTS; PR00205; CADHERIN.
CC PRINTS; PR00011; EGFLAMININ.
CC PRINTS; PR00249; GPCRSECRETIN.
CC SMART; SM00112; CA; 9.
CC SMART; SM00179; EGF_CA; 1.
CC SMART; SM00303; GPS; 1.
CC SMART; SM00008; Hormr; 1.
CC SMART; SM00282; LamG; 2.
CC PROSITE; PS00010; ASX_HYDROXYL; 2.
CC PROSITE; PS00232; CADHERIN_1; 7.
CC PROSITE; PS0268; CADHERIN_2; 9.
CC PROSITE; PS00022; EGF_1; 6.
CC PROSITE; PS01186; EGF_2; 4.
CC PROSITE; PS00026; EGF_3; 6.
CC PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
CC PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
CC PROSITE; PS00227; G_PROTEIN_RECEP_F2_3; 1.
CC PROSITE; PS0261; G_PROTEIN_RECEP_F2_4; 1.
CC PROSITE; PS00221; GPS; 1.
CC PROSITE; PS00025; LAM_G_DOMAIN; 2.
CC PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
KW Developmental protein; Hydroxylation; Signal.
FT SIGNAL 1 31
FT CHAIN 32 2923
FT DOMAIN 32 2380
FT TRANSMEM 2381 2401
FT DOMAIN 2402 2416
FT TRANSMEM 2417 2437
FT DOMAIN 2438 2438

TRANSMEM 2439 2459 3 (POTENTIAL).
DOMAIN 2460 2480 CYTOPLASMIC (POTENTIAL).
TRANSMEM 2481 2501 4 (POTENTIAL).
DOMAIN 2502 2519 EXTRACELLULAR (POTENTIAL).
TRANSMEM 2520 2540 5 (POTENTIAL).
DOMAIN 2541 2560 CYTOPLASMIC (POTENTIAL).
TRANSMEM 2561 2581 6 (POTENTIAL).
DOMAIN 2582 2591 EXTRACELLULAR (POTENTIAL).
TRANSMEM 2592 2612 7 (POTENTIAL).
DOMAIN 2613 2623 CYTOPLASMIC (POTENTIAL).
DOMAIN 182 289 CADHERIN 1.
DOMAIN 290 399 CADHERIN 2.
DOMAIN 400 505 CADHERIN 3.
DOMAIN 506 610 CADHERIN 4.
DOMAIN 611 712 CADHERIN 5.
DOMAIN 713 815 CADHERIN 6.
DOMAIN 816 921 CADHERIN 7.
DOMAIN 922 1023 CADHERIN 8.
DOMAIN 1028 1146 CADHERIN 9.
DOMAIN 1286 1296 EGF-LIKE 1, CALCIUM-BINDING.
DOMAIN 1298 1324 EGF-LIKE 2, CALCIUM-BINDING.
DOMAIN 1328 1366 EGF-LIKE 3, CALCIUM-BINDING.
DOMAIN 1367 1571 LAMININ G-LIKE 1.
DOMAIN 1574 1610 EGF-LIKE 4, CALCIUM-BINDING.
DOMAIN 1614 1791 LAMININ G-LIKE 2.
DOMAIN 1793 1828 EGF-LIKE 5, CALCIUM-BINDING.
DOMAIN 1829 1867 EGF-LIKE 6, CALCIUM-BINDING.
DOMAIN 1883 1922 EGF-LIKE 7, CALCIUM-BINDING.
DOMAIN 1923 1955 EGF-LIKE 8, CALCIUM-BINDING.
DOMAIN 2316 2368 GPS.
DOMAIN 2743 2752 POLY-GLU.
DOMAIN 1232 1243 BY SIMILARITY.
DISULFID 1237 1274 BY SIMILARITY.
DISULFID 1276 1285 BY SIMILARITY.
DISULFID 1292 1303 BY SIMILARITY.
DISULFID 1297 1312 BY SIMILARITY.
DISULFID 1314 1323 BY SIMILARITY.
DISULFID 1332 1343 BY SIMILARITY.
DISULFID 1337 1353 BY SIMILARITY.
DISULFID 1355 1365 BY SIMILARITY.
DISULFID 1578 1589 BY SIMILARITY.
DISULFID 1593 1598 BY SIMILARITY.
DISULFID 1600 1609 BY SIMILARITY.
DISULFID 1757 1808 BY SIMILARITY.
DISULFID 1802 1817 BY SIMILARITY.
DISULFID 1819 1828 BY SIMILARITY.
DISULFID 1832 1843 BY SIMILARITY.
DISULFID 1837 1855 BY SIMILARITY.
DISULFID 1857 1866 BY SIMILARITY.
DISULFID 1887 1899 BY SIMILARITY.
DISULFID 1889 1906 BY SIMILARITY.
DISULFID 1908 1921 BY SIMILARITY.
DISULFID 1924 1936 BY SIMILARITY.
DISULFID 1926 1943 BY SIMILARITY.
DISULFID 1945 1954 BY SIMILARITY.
MOD_RES 1591 1591 HYDROXYLATION (POTENTIAL).
MOD_RES 1591 1591 HYDROXYLATION (POTENTIAL).
CARBOHYD 486 486 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 701 701 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 1036 1036 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 1076 1076 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 1182 1182 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 1212 1212 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 1501 1501 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 1565 1565 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 1741 1741 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 1837 1827 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 1900 1900 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 2024 2024 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 2043 2043 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 2051 2061 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 2323 2323 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2345 2345 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 2923 AA; 317447 MW; 382757D315158ED8 CRC64;
Query Match 13.5%; Score 211.5; DB 1; Length 2923;
Best Local Similarity 23.3%; Pred No. 3.4e-07;
Matches 81; Conservative 43; Mismatches 132; Indels 91; Gaps 15;
QY 1 CERQPCQHCATCMPAGEYE-----FQCLCRDGFKGLCEHENP 39
DB 1232 CLREPCENTMRCVSLRFDSSAFFIASSVLPFRPHVPGGLRCPCPGFTGDCYCEYDL 1291
QY 40 COLREPLHGGTCQ---GTRCLCLPGFSGPRCCQG--SGH---GIASSDWH---LEGS 86
DB 1292 CVSR-PCGPHGCRSRREGGTCCLCRDGYTGHECEVSARSGRCTPGVCXNGGTCVNLVGG 1350
QY 87 GGNDAPE-QQYGAIFYHDDGFLAPPGHVP--SRSLPE-VPETIELEVRTSTAGLLWQGV 142
DB 1351 FXDCDPSGDFEXPCYQVITRSPFAHSFIFRGLRQRFFHTLALSFPATKXRDGLLYN--- 1407
QY 143 VGEAGQKDFISLGLQDGHVFRYQLGSGEARLVSEDP--INDGEWHRV----- 189
DB 1408 -GRFNEKHDFALEVIQSQVLTFSAGESTTTSPFVPGVSDGGQWHTVQLKYKPLLG 1466
QY 190 -TALREG-----RGSIQVDGEELVSGRSPGNVAVNAKSVYIG 228
DB 1467 QTGLPQGPSEQRVAVVTVDCDGTVALRFGVLGNSCAAQGTQGSKSLDLTGPLLLG 1526
QY 229 GAPDYATLTGGRFSSGIGTCVKNLVLSAR-----PGAP 262
DB 1527 GVFDLPESPFPVRNRQ-FVGCNRNLQVDSRHIDMADFIANNGTVPQCP 1572
RESULT 14
CLAR2_RAT
ID CLAR2_RAT STANDARD; PRT; 2144 AA.
AC Q9QTF2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cadherin EGF LAG seven-pass G-type receptor 2 (Multiple epidermal
growth factor-like domains 3) (Fragment).
GN CELSR2 OR VEGF3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=98360089; PubMed=983030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple
EGF-like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell
signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed in the brain. High expression in
cerebellum and olfactory bulb. Weaker expression in cerebral
cortex, hippocampus and brain stem.
CC -1- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
CC -1- SIMILARITY: Contains 8 cadherin domains.
CC -1- SIMILARITY: Contains 4 EGF-like domains.
CC -1- SIMILARITY: Contains 2 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 laminin EGF-like domain.
CC -1- SIMILARITY: Contains 1 GPS domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)


```

or send an email to license@isb-sib.ch).
-----
CC CC      EMBL; AB011529; BA88687.1; -.
DR DR      HSPF; P00740; IEDM.
DR DR      GO; GO:0016021; C: integral to membrane; ISS.
DR DR      GO; GO:0004930; F: G-protein coupled receptor activity; ISS.
DR DR      GO; GO:0007186; P: G-protein coupled receptor protein signalin. .; ISS.
DR DR      InterPro; IPR000152; Asx_hydroxyl_S.
DR DR      InterPro; IPR002126; Cadherin.
DR DR      InterPro; IPR008985; Cna_like_lec_gl.
DR DR      InterPro; IPR000742; EGF_2.
DR DR      InterPro; IPR001881; EGF_Ca.
DR DR      InterPro; IPR006209; EGF_Like.
DR DR      InterPro; IPR000832; GPCR secretin.
DR DR      InterPro; IPR001879; hormn_receptor.
DR DR      InterPro; IPR002049; Laminin_EGF.
DR DR      InterPro; IPR001791; Laminin_G.
DR DR      InterPro; IPR000203; PKD_cys_rich.
DR DR      Pfam; PF00002; 7tm_2; 1.
DR DR      Pfam; PF00028; cadherin; 3.
DR DR      Pfam; PF00008; EGF; 5.
DR DR      Pfam; PF01825; GPS; 1.
DR DR      Pfam; PF02793; HRM; 1.
DR DR      Pfam; PF00053; laminin_EGF; 1.
DR DR      Pfam; PF00054; laminin_G; 1.
DR DR      PRINTS; PR00205; CADHERIN.
DR DR      PRINTS; PR00011; EGF_LAMININ.
DR DR      PRINTS; PR00249; GPCRSECRETIN.
DR DR      SMART; SM00112; CA; 3.
DR DR      SMART; SM00179; EGF_CA; 1.
DR DR      SMART; SM00303; GPS; 1.
DR DR      SMART; SM00008; Hormr; 1.
DR DR      SMART; SM00282; Lamg; 2.
DR DR      PROSITE; PS00010; ASX_HYDROXYL; 2.
DR DR      PROSITE; PS00232; CADHERIN_1; 3.
DR DR      PROSITE; PS00268; CADHERIN_2; 4.
DR DR      PROSITE; PS00022; EGF_1; 6.
DR DR      PROSITE; PS01186; EGF_2; 4.
DR DR      PROSITE; PS00026; EGF_3; 6.
DR DR      PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR DR      PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
DR DR      PROSITE; PS00227; G_PROTEIN_RECEP_F2_3; 1.
DR DR      PROSITE; PS00261; G_PROTEIN_RECEP_F2_4; 1.
DR DR      PROSITE; PS00221; GPS; 1.
DR DR      PROSITE; PS50025; LAM_G_DOMAIN; 2.
DR DR      PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KH KH      G-protein coupled receptor; Transmembrane; Glycoprotein;
KW KW      EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
KW KW      Developmental protein; Hydroxylation.
FT FT      NON_TER 1
FT FT      DOMAIN 1 1605 EXTRACELLULAR (POTENTIAL).
FT FT      TRANSMEM 1606 1 (POTENTIAL).
FT FT      DOMAIN 1626 2 (POTENTIAL).
FT FT      TRANSMEM 1627 1641 CYTOPLASMIC (POTENTIAL).
FT FT      TRANSMEM 1642 1662 2 (POTENTIAL).
FT FT      DOMAIN 1663 1663 EXTRACELLULAR (POTENTIAL).
FT FT      TRANSMEM 1664 1684 3 (POTENTIAL).
FT FT      DOMAIN 1685 1705 CYTOPLASMIC (POTENTIAL).
FT FT      TRANSMEM 1706 1726 4 (POTENTIAL).
FT FT      DOMAIN 1727 1744 EXTRACELLULAR (POTENTIAL).
FT FT      TRANSMEM 1745 1765 5 (POTENTIAL).
FT FT      DOMAIN 1766 1789 CYTOPLASMIC (POTENTIAL).
FT FT      TRANSMEM 1790 1810 6 (POTENTIAL).
FT FT      DOMAIN 1811 1816 EXTRACELLULAR (POTENTIAL).
FT FT      TRANSMEM 1817 1837 7 (POTENTIAL).
FT FT      DOMAIN 1838 2144 CYTOPLASMIC (POTENTIAL).
FT FT      TRANSMEM <1 40 CADHERIN_1.
FT FT      DOMAIN 41 146 CADHERIN_2.
FT FT      TRANSMEM 147 248 CADHERIN_3.
FT FT      DOMAIN 253 371 CADHERIN_4.
FT FT      TRANSMEM 453 511 EGF-LIKE 1.
FT FT      DOMAIN 513 549 EGF-LIKE 2.
FT FT      TRANSMEM 553 591 EGF-LIKE 3.
FT FT      DOMAIN 592 796 LAMININ G-LIKE 1.

```

```

FT DOMAIN 799 835
FT DOMAIN 839 1016
FT DOMAIN 1018 1053
FT DOMAIN 1054 1092
FT DOMAIN 1108 1147
FT DOMAIN 1148 1180
FT DOMAIN 1541 1593
FT DOMAIN 1968 1973
FT DISULFID 457 468
FT DISULFID 462 499
FT DISULFID 501 510
FT DISULFID 517 528
FT DISULFID 522 537
FT DISULFID 539 548
FT DISULFID 557 568
FT DISULFID 562 578
FT DISULFID 580 590
FT DISULFID 803 814
FT DISULFID 808 823
FT DISULFID 825 834
FT DISULFID 1022 1033
FT DISULFID 1027 1042
FT DISULFID 1044 1053
FT DISULFID 1057 1068
FT DISULFID 1062 1080
FT DISULFID 1082 1091
FT DISULFID 1112 1124
FT DISULFID 1114 1131
FT DISULFID 1133 1146
FT DISULFID 1149 1161
FT DISULFID 1151 1168
FT DISULFID 1170 1179
FT MOD_RES 816 816
FT MOD_RES 1035 1035
FT CARBOHYD 261 261
FT CARBOHYD 301 301
FT CARBOHYD 407 407
FT CARBOHYD 437 437
FT CARBOHYD 726 726
FT CARBOHYD 790 790
FT CARBOHYD 966 966
FT CARBOHYD 1052 1052
FT CARBOHYD 1125 1125
FT CARBOHYD 1249 1249
FT CARBOHYD 1268 1268
FT CARBOHYD 1286 1286
FT CARBOHYD 1548 1548
FT CARBOHYD 1570 1570
FT CARBOHYD 2144 AA; 233480 MW; 6EA898C1BA655ECA CRC64;
SQ SEQUENCE 2144 AA; 233480 MW; 6EA898C1BA655ECA CRC64;

Query Match 13.3%; Score 207.5; DB 1; Length 2144;
Best Local Similarity 23.6%; Pred. No. 4,7e-07;
Matches 82; Conservative 42; Mismatches 132; Indels 91; Gaps 16;

QY 1 CERPCOHGATCMPAGEYE-----FOCLCRDGGKGLCEHEENP 39
Db 457 CLREPCENYRCVSLRFDSSAFIASSVLFPIHPVGLRCRCPPGTGDCETEVLD 516
QY 40 COLRBPCLHGGTCQ----GTRCLCLPFGSGPQQG--SGH---GIAESDWH----LEGS 86
Db 517 CYSR-PCGPHGCHCRSREGVTCICRDGYTGCHCEVSARSGRCTPGVCKNGGTCVNLVGG 575
QY 87 GGNDAF-GQYGAYFHDHGGFLAPPGHYF--SRSLPE-VPEITIELEVSTASGULLWQGYE 142
Db 576 FKDCPSGDFEKFPCQVTRSPFARSFIFRGLRQRFHFTLASFATKRDGGLLYN--- 632
QY 143 VGEAGQGGKDFISLGLQDGHVLFYQLGSGEARLVSEDP--INDGEWHRV----- 189
Db 633 -GRFNEKHDFVALEVIQEQVQLTFSAGESTTTVPVPGVSDGQWHTVOLKYNKPLLG 691
QY 190 -TALRG---RRGSIQVDEE-----LVSGRSPPGNVAVNAKGVYTG 228
Db 692 QTGLPQSPSEQKVAVVSVVDGCDTGVALRFGAMLGNYSACAAQQTGGSKSLDTGFLILG 751

```

EGF-LIKE 4, CALCIUM-BINDING.
LAMININ G-LIKE 2.

EGF-LIKE 5, CALCIUM-BINDING.
EGF-LIKE 6, CALCIUM-BINDING.
EGF-LIKE 7, CALCIUM-BINDING.
EGF-LIKE 8, CALCIUM-BINDING.
GPS.

POLY-GLU.
BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

QY 229 GAPDVTATLTCGFFSGITGCVKXNVLSHAR-----PCAP 262
 DB 752 GVFDLPESFVVRMRH-FVCGMKQLQVDSRHMADFIANNCTVPCP 797

RESULT 15
 LMA2 MOUSE
 ID LMA2 MOUSE STANDARD; PRT; 3106 AA.
 AC Q60675; Q05003; Q64061;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Laminin alpha-2 chain precursor (Laminin M chain) (Merossin heavy chain).
 GN LMA2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Embryo, and Heart;
 RX MEDLINE=95316259; PubMed=7795883;
 RA Bernier S.M., Utani A., Sugiyama S., Doi T., Polistina C., Yamada Y.;
 RT "Cloning and expression of laminin alpha 2 chain (M-chain) in the mouse."
 RL Matrix Biol. 14:447-455(1995).
 RN [2]
 RP SEQUENCE OF 2162-2279 FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Thymus;
 RX MEDLINE=93346725; PubMed=8345183;
 RA Chang A.C., Wadsworth S., Colligan J.E.;
 RT "Expression of merossin in the thymus and its interaction with thymocytes."
 RL J. Immunol. 151:1789-1801(1993).
 RN [3]
 RP SEQUENCE OF 64-281 FROM N.A.
 RX MEDLINE=95179178; PubMed=7874173;
 RA Xu H., Wu X.R., Wewer U.M., Engvall E.;
 RT "Murine muscular dystrophy caused by a mutation in the laminin alpha 2 (Lama2) gene."
 RL Nat. Genet. 8:297-302(1994).
 RN [4]
 RP SEQUENCE OF 20-25.
 RX MEDLINE=21818471; PubMed=11829758;
 RA Garbe J.H., Gohring W., Mann K., Timpl R., Sasaki T.;
 RT "Complete sequence, recombinant analysis and binding to laminins and sulphated ligands of the N-terminal domains of laminin alpha3B and alpha5 chains."
 RL Biochem. J. 362:213-221(2002).
 RN [5]
 RP SEQUENCE OF 20-25.
 RX MEDLINE=21818471; PubMed=11829758;
 RA Taltz J.F., Andac Z., Gohring W., Brancaccio A., Timpl R.;
 RT "Binding of the G domains of laminin alpha1 and alpha2 chains and perlecan to heparin, sulfatides, alpha-dystroglycan and several extracellular matrix proteins."
 RL EMBO J. 18:863-870(1999).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.35 ANGSTROMS) OF 2932-3106.
 RX MEDLINE=20085745; PubMed=10619025;
 RA Hohenester E., Tisi D., Taltz J.F., Timpl R.;
 RT "The crystal structure of a laminin G-like module reveals the molecular basis of alpha-dystroglycan binding to laminins, perlecan, and agrin."
 RL Cell 114:783-792(1999).
 CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin is thought to mediate the attachment, migration and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components.
 CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three

different polypeptide chains (alpha, beta, gamma), which are bound to each other by disulfide bonds into a cross-shaped molecule comprising one long and three short arms with globules at each end. The alpha-2 chain is a subunit of laminin-2 (Merossin) and laminin-4 (S-merossin). Interacts with FBLN1, FBLN2 and NID2.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: Found in the basement membranes (major component).
 CC -1- DOMAIN: The alpha-helical domains I and II are thought to interact with other laminin chains to form a coiled coil structure.
 CC -1- DOMAIN: Domains VI, IV and G are globular.
 CC -1- DISEASE: Defects in LAMA2 are a cause of murine muscular dystrophy (dy2J).
 CC -1- SIMILARITY: Contains 1 laminin N-terminal domain.
 CC -1- SIMILARITY: Contains 17 laminin EGF-like domains.
 CC -1- SIMILARITY: Contains 2 laminin IV domains.
 CC -1- SIMILARITY: Contains 5 laminin G-like domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC EMBL; U12147; AAC52165.1; -
 DR EMBL; X69869; CAA49502.1; -
 DR EMBL; S75315; AAB33573.1; -
 DR FNR; I49077; S53868.
 DR PDB; 1QU0; 03-DEC-99.
 DR PDB; 1DYK; 04-FEB-01.
 DR MGD; MGI:99912; Lama2.
 DR GO; GO:0005604; C:basement membrane; IDA.
 DR InterPro; IPR008985; CoA like lec_g1.
 DR InterPro; IPR006209; EGF like_bind.
 DR InterPro; IPR008979; Gal_bind.
 DR InterPro; IPR000034; Laminin B.
 DR InterPro; IPR002049; Laminin EGF.
 DR InterPro; IPR001791; Laminin G.
 DR InterPro; IPR008211; LamNT.
 DR Pfam; PF00052; laminin B; 2.
 DR Pfam; PF00053; laminin_EGF; 14.
 DR Pfam; PF00054; laminin_G; 5.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR ProDom; PD003031; Laminin B; 1.
 DR SMART; SM00180; EGF Lam; 14.
 DR SMART; SM00281; LamB; 2.
 DR SMART; SM00282; LamG; 5.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00022; EGF_1; 11.
 DR PROSITE; PS01186; EGF_2; 3.
 DR PROSITE; PS01248; LAMININ TYPE EGF; 14.
 DR PROSITE; PS00025; LAM G DOMAIN; 5.
 DR GlycoProtex; Basement membrane; Extracellular matrix; Coiled coil; Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.
 FT SIGNAL 1 19
 FT CHAIN 20 3106 LAMININ ALPHA-2 CHAIN.
 FT DOMAIN 20 282 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 283 339 LAMININ EGF-LIKE 1.
 FT DOMAIN 340 409 LAMININ EGF-LIKE 2.
 FT DOMAIN 410 464 LAMININ EGF-LIKE 3.
 FT DOMAIN 465 513 LAMININ EGF-LIKE 4.
 FT DOMAIN 514 523 LAMININ EGF-LIKE 5 (N-TERMINAL).
 FT DOMAIN 524 719 LAMININ DOMAIN IV 1 (DOMAIN IV B).
 FT DOMAIN 720 752 LAMININ EGF-LIKE 5 (C-TERMINAL).
 FT DOMAIN 753 802 LAMININ EGF-LIKE 6.
 FT DOMAIN 803 860 LAMININ EGF-LIKE 7.
 FT DOMAIN 861 913 LAMININ EGF-LIKE 8.
 FT DOMAIN 914 962 LAMININ EGF-LIKE 9.
 FT DOMAIN 963 1009 LAMININ EGF-LIKE 10.
 FT DOMAIN 1010 1055 LAMININ EGF-LIKE 11.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:15:08 ; Search time 24.8707 Seconds

(without alignments)
3602.917 Million cell updates/sec

Title: US-10-006-011A-9

Perfect score: 1566

Sequence: 1 CERQPCQHGATCMPAGEYEF.....QPLDLQRAQAGANTPCPS 284

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriaph:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	357.5	22.8	2026	4	O00468
2	349	22.3	925	5	Q9U7E8
3	346.5	22.1	488	4	Q96IC1
4	343	21.9	463	4	Q8NAL2
5	343	21.9	775	4	Q8N7Y0
6	325	20.8	295	11	Q80WX4
7	325	20.8	1009	11	Q8BGP3
8	304	19.4	283	11	Q8QV56
9	304	19.4	283	11	Q99KT4
10	304	19.4	350	11	Q8K326
11	295	18.8	478	5	Q86SD6
12	292	18.6	1361	5	Q9V7I4
13	292	18.6	1361	5	Q9NGV2
14	292	18.6	1534	4	Q75093
15	292	18.6	1618	4	Q9U1L7
16	283.5	18.1	1531	11	O88279

17 281 17.9 708 5 Q9NFS9

18 281 17.9 1035 5 Q9NEG1

19 281 17.9 4117 5 Q8IRV9

20 281 17.9 4179 5 Q9W4Y4

21 281 17.9 4223 5 Q8MPN3

22 281 17.9 4228 5 Q8IRV8

23 280.5 17.9 738 13 Q90Z45

24 280.5 17.9 1458 11 Q8CJG8

25 280.5 17.9 1531 11 Q9WVB5

26 280.5 17.9 1557 11 Q80TR4

27 279.5 17.8 68 11 Q08591

28 278.5 17.8 152 4 Q8N124

29 276 17.6 152 4 Q8N197

30 274.5 17.5 1530 11 Q9WUG5

31 270.5 17.3 1474 11 Q8CJG5

32 265 16.9 181 4 Q8N4J5

33 264 16.9 4307 5 Q19319

34 258.5 16.5 1529 13 Q7ZXI2

35 250.5 16.0 1039 5 Q9VQB1

36 249 15.9 1410 5 Q20204

37 246.5 15.7 1515 13 Q8DE37

38 246 15.7 1523 11 Q88280

39 244.5 15.6 1512 13 Q9DE36

40 244 15.6 1530 13 Q90WZ3

41 243.5 15.5 1025 11 Q9Z166

42 243.5 15.5 1521 11 Q9R1B9

43 241.5 15.4 783 13 Q90XG2

44 241.5 15.4 950 13 Q90Z44

45 241 15.4 1523 4 Q75094

ALIGNMENTS

RESULT 1

O00468 PRELIMINARY; PRT; 2026 AA.

ID O00468

AC O00468; 01-JUL-1997 (TREMBLrel. 04, Created)

DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE AGRIN precursor (Fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Groffen A.J.A., Buskens C.A.F., Van Kuppevelt T.H.M.S.M.,

RA Veerkamp J.H., Momms L.A.H., van den Heuvel L.P.W.J.;

RL Eur. J. Biochem. 0:0-0(1998).

RN [2]

RP SEQUENCE OF 1-153 FROM N.A.

RX MEDLINE=96224170; PubMed=8617505;

RA Lennon G., Auffray C., Polymeropoulos M., Soares M.B.;

RT "The I.M.A.G.E. Consortium: an integrated molecular analysis of

RT genomes and their expression.";

RL Genomics 33:151-152(1996).

RN [3]

RP SEQUENCE OF 1-153 FROM N.A.

RA Denzer A.J., Brandenberger R., Gesemann M., Chiquet M., Ruegg M.A.;

RL J. Cell Biol. 0:0-0(0).

DR EMBL; AF016903; AAC39776.1; -.

DR EMBL; U84406; AAB52917.1; -.

DR HSSP; P00740; IEDM.

DR Genew; HGNC:329; AGRN.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR InterPro; IPR004850; Agrin NtA.

DR InterPro; IPR006209; EGF like.

DR InterPro; IPR003645; FcN.

DR InterPro; IPR002350; Kazal.

DR InterPro; IPR002049; Laminin EGF.

DR InterPro; IPR001791; Laminin G.

```

DR InterPro; IPR000082; SEA domain.
DR InterPro; IPR008993; TIMP-like.
DR InterPro; IPR001455; UPP0033.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00050; Kazal; 9.
DR Pfam; PF00053; laminin_EGF; 2.
DR Pfam; PF00054; laminin_G; 3.
DR Pfam; PF03146; NEA; 1.
DR Pfam; PF01390; SEA; 1.
DR PRINTS; PRO0011; EGF_LAMININ.
DR SMART; SM00180; EGF_Lam; 2.
DR SMART; SM00274; FOLN; 5.
DR SMART; SM00280; KAZAL; 9.
DR SMART; SM00282; LamG; 3.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; 6.
DR PROSITE; EGF_2; 1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
DR PROSITE; PS00024; SEA; 1.
DR PROSITE; PS01148; UPP0033; 1.
KW EGF-like domain; Laminin EGF-like domain; Signal.
FT NON_TER 1
FT SIGNAL <1 10 POTENTIAL.
FT CHAIN 11 2026 AGRIN.
SQ SEQUENCE 2036 AA; 212881 MW; 4AB0EE710CD4B8EF CRC64;

Query Match 22.8%; Score 357.5; DB 4; Length 2026;
Best Local Similarity 35.4%; Pred. No. 4.7e-19;
Matches 91; Conservative 29; Mismatches 110; Indels 27; Gaps 8;

QY 1 CERQPCQHGATCMPAGEYFQCLRDGKGLDCEHENPCQLRPFCLHGGTCQ-----GT 55
DB 1534 CLPNPCHGAPCONLEAGRFHCQCPGRVGTCADEKSPCQ-PNPGHGAAPCVLPESGA 1592

QY 56 RCLCLPGFSGPRCCQSGHGIAESDWHLEGGGNDAPQCYGAYFHDGFLAPG-HVFSR 114
DB 1593 QCECPZLREGTFCQAS-----GQGGSPFLADPFGSHLELRLGHTFAR 1637

QY 115 SLPEVPETIELEVRTSTASGLLMQGVVEGAGQKDFISLGLQDGHVFRYQLGSGEAR 174
DB 1638 DLGE-KMALEVFLARGPSGLLYNGKTD--GKG-DFVSLARDRLFRYDLGKGA 1693

QY 175 LVSEDPINDGEHVRTALREGRSGIQVDGELVSGRSPGNVAVNAKGSVYICGAPDVA 234
DB 1694 IRSREPVTLGATWRVSLERNKGALRYGDDGPRVLGESPVHTVTLNKLPELYVGAPDFS 1753

QY 235 TLT-GGRFSSGTTGCVK 250
DB 1754 KLARAANVSSGFDGAIQ 1770

RESULT 2
Q907E8 PRELIMINARY; PRT; 925 AA.
AC Q907E8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE UNC-52/perlecan (Fragment).
GN UNC-52.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N2;
RX MEDLINE=99443780; PubMed=10512861;
RA Mullen G.P., Rogalski T.M., Bush J.A., Gortj P.R., Moerman D.G.;
RT "Complex patterns of alternative splicing mediate the spatial and
RT temporal distribution of Perlecan/UNC-52 in caenorhabditis elegans.";
RL Mol. Biol. Cell 10:3205-3221(1999).

```

```

DR EMBL; AFL32883; AAD25092.1; -.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00008; EGF; 3.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00054; laminin_G; 2.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00408; IGC2; 1.
DR SMART; SM00282; LamG; 3.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS00835; IGLIKE; 1.
DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
KW EGF-like domain; Immunoglobulin domain.
FT NON_TER 1
SQ SEQUENCE 925 AA; 102174 MW; 474BBSF045D67E0B CRC64;

Query Match 22.3%; Score 349; DB 5; Length 925;
Best Local Similarity 31.2%; Pred. No. 8.1e-19;
Matches 81; Conservative 41; Mismatches 102; Indels 36; Gaps 6;

QY 30 GDLCEHEE--NPCQLRPECPLHGGTCQGT-----CLCLPGFSGPRCCQSGHGIAESDWH 82
DB 679 GDVYSTQEPNNIC-ANSTCGMNGQCVPRNMTHYTCEKLYYDGTCSLKF----- 727

QY 83 LEGSGGNDAPQCYGAYFHDGFLAPGHVPSRSLPEVETIELEVRTSTASGLLMQGV 142
DB 728 -----PIBHAARFDGDAFIELSSDEFPHTSEKDEIVAFKPTQOQNGVLLWQG-Q 777

QY 143 VGEAGQKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGEHVRTALREGRSGIQV 202
DB 778 RPTVQQMEDYISVGVNGHLHFSYELGGAAHLISEERVDDCKEHSVRFERKREGQMRI 837

QY 203 DEELVSGRSPGNVAVNAKGSVYTGGAADVATLTCGRFSSGTCVKVLVLSARPAP 262
DB 838 DNYREVDGSGTILANLVNDGNIFVGVGVDISKATGGLFSNNFVGCIADVELNGVK---- 893

QY 263 PPQDLQLHRAQAGANTRPC 282
DB 894 ----LDLMATAIDGNVKFC 909

RESULT 3
Q96IC1 PRELIMINARY; PRT; 488 AA.
AC Q96IC1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007649; AAH07649.1; -.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00008; EGF; 3.
DR Pfam; PF00054; laminin_G; 2.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00282; LamG; 2.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 1.

```

```
DR PROSITE; PS00025; LAM G DOMAIN; 2.  
KW Hypothetical protein; EGF-like domain.  
FT NON_TER 1  
SQ SEQUENCE 488 AA; 51764 MW; A732B99CC80E772 CRC64;  
  
Query Match 22.1%; Score 346.5; DB 4; Length 488;  
Best Local Similarity 35.3%; Pred. No. 5.6e-19;  
Matches 89; Conservative 29; Mismatches 107; Indels 27; Gaps 8;  
  
QY 5 CQHATCMFAGEVEROCLEKDFKDLCEHENPCQLREPCPLHGTCQ-----GTRCLCL 60  
DB 1 CHGGAPCQNLGRFCPCPPGRVGTCADEKSPQ-PNPGGAAPCRVLPBGGAQCPCP 59  
  
QY 61 PGFSGPRCQCGSGHGAESDWHLESGGNDAPGQYCAFYHDDGFLAFFG-HVFSRSLPEV 119  
DB 60 LGREGFTFCQAS-----GQDGGPPLADFNFGPSHLELRLGHTFARDLGE- 103  
  
QY 120 PETIELEVTSTASGLLLWQGVGEAGQCKDFISLGLDQCHLVFRYQLGSGEARLVSED 179  
DB 104 KMALEVFLARGPSGLLYNGQKTD--GKG-DFVSLALDRLEFRYDLGKGAIVRSRE 160  
  
QY 180 PINDGEWHRTALREGRRGSIQVDGSELVSGRSPGNVAVNAKSGVYIGGAPDVATLT-G 238  
DB 161 PVTLGATVTRVLERNGKALRGVDCGPRVLGSPVPHVTNLKEPLYVGADPFSKLARA 220  
  
QY 239 GRPSSGITGVCK 250  
DB 221 AAVSSGDFGAIQ 232  
  
RESULT 4  
Q8NAL2 PRELIMINARY; PRT; 463 AA.  
AC Q8NAL2;  
DT 01-OCT-2002 (T-EMBLrel. 22, Created)  
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
DE Hypothetical protein FLJ35160.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,  
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,  
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,  
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,  
RA "NEDO human cDNA sequencing project."  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK092479; BAC03900.1;  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR InterPro; IPR001881; EGF_CA.  
DR InterPro; IPR006209; EGF-like.  
DR InterPro; IPR006210; IEGF.  
DR InterPro; IPR001791; Laminin_G.  
DR Pfam; PF00008; EGF; 2.  
DR Pfam; PF00054; laminin_G; 3.  
DR SMART; SM00181; EGF; 3.  
DR SMART; SM00179; EGF_CA; 2.  
DR SMART; SM00282; LamG; 2.  
DR PROSITE; PS00022; LamG; 3.  
DR PROSITE; PS01186; EGF 1; 2.  
DR PROSITE; PS01186; EGF 2; 2.  
DR PROSITE; PS00025; LAM G DOMAIN; 3.  
KW Hypothetical protein; EGF-like domain.  
SQ SEQUENCE 463 AA; 50635 MW; F688BF2714D5D08C CRC64;  
  
Query Match 21.9%; Score 343; DB 4; Length 463;  
Best Local Similarity 33.5%; Pred. No. 9.9e-19;  
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;  
  
QY 34 EHENPCQLREPCPLHGTC-----QTRCLCLPFGSGPRCQCGSGHGAESDWHLESGGN 89  
DB 236 ENAAHPC-VRAPCAHGSGCRPRXEGYDCDCPLGFEGLHCOKAIIETI----- 281  
  
QY 90 DAP-----GQYCAFYHDDGFLAFFGHPVFSRSLPEVPTIELEVTSTASGLLLWQGVGE 145  
DB 282 EIPQFGRSYLTNDPDLKRVSG-----SRV-----NVFMRFKITAKDGLLLWRG--DSP 330  
  
QY 146 AGQCKDFISLGLDQCHLVFRYQLGSGEARLVSEDPINDGEWHRTALREGRRGSIQVDGE 205  
DB 331 MRPNDFISLGLRDGALVFSYNLGSGVASIWNVNGSFNDGHRVKAQRDQSGSKITVDDY 390  
  
QY 206 ELVSGRSPGNVAVNAKSGVYIGGAPDVATLTGFRSSGITGVCKNLVLSARPAPPPQ 265  
DB 391 GARTGSPGMRQLNTGALYVGGMKETALHTNRQYMRGLVGCISHFTLT-----DY 443  
  
QY 266 PLDLQHEAQAQAGANTRPC 282  
DB 444 HISLVEDAVDVGKINTC 460  
  
RESULT 5  
Q8N7Y0 PRELIMINARY; PRT; 775 AA.  
AC Q8N7Y0;  
DT 01-OCT-2002 (T-EMBLrel. 22, Created)  
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
DE Hypothetical protein FLJ40230.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,  
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,  
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,  
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,  
RA "NEDO human cDNA sequencing project."  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK097549; BAC05096.1;  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR InterPro; IPR008985; ConA_like lec_g.  
DR InterPro; IPR001881; EGF_CA.  
DR InterPro; IPR006209; EGF-like.  
DR InterPro; IPR006210; IEGF.  
DR InterPro; IPR001791; Laminin_G.  
DR Pfam; PF00008; EGF; 2.  
DR Pfam; PF00054; laminin_G; 3.  
DR SMART; SM00181; EGF; 3.  
DR SMART; SM00179; EGF_CA; 2.  
DR SMART; SM00282; LamG; 3.  
DR PROSITE; PS00022; LamG; 3.  
DR PROSITE; PS01186; EGF 1; 2.  
DR PROSITE; PS01186; EGF 2; 2.  
DR PROSITE; PS00025; LAM G DOMAIN; 3.  
KW Hypothetical protein; EGF-like domain.  
SQ SEQUENCE 775 AA; 84800 MW; 61A81294F0204ED2 CRC64;  
  
Query Match 21.9%; Score 343; DB 4; Length 775;  
Best Local Similarity 33.5%; Pred. No. 1.9e-18;  
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;  
  
QY 34 EHENPCQLREPCPLHGTC-----QTRCLCLPFGSGPRCQCGSGHGAESDWHLESGGN 89  
DB 548 ENAAHPC-VRAPCAHGSGCRPRXEGYDCDCPLGFEGLHCOKAIIETI----- 593  
  
QY 90 DAP-----GQYCAFYHDDGFLAFFGHPVFSRSLPEVPTIELEVTSTASGLLLWQGVGE 145
```